

УНИВЕРЗИТЕТ У БЕОГРАДУ

ФИЛОЛОШКИ ФАКУЛТЕТ

Александра С. Тртовац

**ДЕСКРИПТОРИ МЕТАПОДАТАКА И
ДЕСКРИПТОРИ САДРЖАЈА У ПРОНАЛАЖЕЊУ
ИНФОРМАЦИЈА У ДИГИТАЛНИМ
БИБЛИОТЕКАМА**

докторска дисертација

Београд, 2016.

UNIVERSITY OF BELGRADE

FACULTY OF PHILOLOGY

Aleksandra S. Trtovac

**METADATA DESCRIPTORS AND CONTENT
DESCRIPTORS FOR INFORMATION RETRIEVAL
IN DIGITAL LIBRARIES**

doctoral dissertation

Belgrade, 2016.

Ментор:

др Цветана Крстев, редовни професор, Универзитет у Београду, Филолошки факултет

Чланови комисије:

др Александра Вранеш, редовни професор, Универзитет у Београду, Филолошки факултет

др Ранка Станковић, ванредни професор, Универзитет у Београду, Рударско-геолошки факултет

др Рајна Драгићевић, редовни професор, Универзитет у Београду, Филолошки факултет

др Душко Витас, редовни професор, Универзитет у Београду, Математички факултет

Датум одбране: _____

Захвална сам свом ментору, проф. др Цветани Крстев, на свим саветима, сугестијама и безрезервној подршци током израде ове докторске дисертације.

Захваљујем се и својим драгим колегиницама и колегама из Универзитетске библиотеке „Светозар Марковић“ који су ми небројено пута изашли у сусрет током докторских студија и рада на дисертацији.

Посвећено Емилу и Селману

Дескриптори метаподатака и дескриптори садржаја у проналажењу информација у дигиталним библиотекама

Резиме

У овој докторској дисертацији испитују се перформансе проналажења информација у дигиталним библиотекама коришћењем различитих метода – дескриптора метаподатака и дескриптора садржаја.

Развој библиотека од традиционалне, преко електронске до дигиталне довео је и до промене метода за проналажење информација. Како је временом дошло до огромног пораста количине информација, тако су се променили и њихови носиоци, па је све већа количина дигитализоване аналогне грађе, али и изворно дигиталних објеката. Ови објекти имају своје особености које су значајне за проналажење информација.

Дигиталне објекте је, у циљу проналажења информација, могуће описати дескрипторима метаподатака различитих врста и у различитим форматима зависно од система у којима су ти објекти похрањени. Такође, дигитални објекти, могу бити опремљени и дескрипторима садржаја што подразумева коришћење напредних технологија за пречишћавање текста (оптичко препознавање карактера, оптичку сегментацију чланака, препознавање именованих ентитета), као и лексичке ресурсе и алате који доприносе бољем и прецизнијем проналажењу информација у дигиталним библиотекама.

Анализа претраге дигиталних објеката и проналажење информација о њима показује да без обзира на све функционалности које нуде различити формати за унос метаподатака, као и везе дигиталног објекта са семантичким вебом, претрага преко метаподатака не даје увек добре резултате у смислу прецизности. С друге стране, претрага преко комплетног текста коришћењем дескриптора садржаја нуди неупоредиво боље резултате нарочито у оним дигиталним колекцијама које су опремљене квалитетним лексичким ресурсима, у првом реду морфолошким речницима.

У циљу побољшања проналажења информација у дигиталним библиотекама из области библиотекарства и информатике надградили смо и вишечланим речима допунили лексичке ресурсе српског језика – термилошки речник из поменуте области, као и општи речник српског језика при чему су поштовани стандарди за израду електронских термилошких речника и термилошких база података. У процесу допуне поменутих лексичких ресурса анализирани су текстови у доменском корпусу, испитана функционалност постојећих коначних трансдуктора у екстракцији

вишечланих лема из доменског корпуса, проверена исправност препознатих лема и извршена тростепена евалуације екстрахованих вишечланих речи. Практични рад резултирао је допуном термилошког речника са скоро 600 нових вишечланих термина, а надограђен је и општи електронски речник вишечланих речи српског језика. Допуњени речници примењени у екстракцији термина на новом корпусу текстова су дали веома добре резултате.

На основу спроведеног истраживања и анализа закључили смо да напредне методе помажу у квалитетнијем проналажењу информација. То је потврђено и упоређивањем ефикасности и прецизности проналажења информација уз примену напредних метода у односу на класичне методе. Такође, можемо рећи да проналажење информација помоћу дескриптора садржаја даје немерљиво боље резултате у односу на проналажење информација које се ослања само на дескрипторе метаподатака. Неопходно је истаћи важност коришћења лексичких ресурса међу којима посебно место заузимају морфолошки речници што је од изузетног значаја за језике богате флексије какав је српски језик.

Кључне речи: проналажење информација, дигиталне библиотеке, дескриптори метаподатака, дескриптори садржаја, лексички ресурси, електронски речници, термилошки речници, српски језик

Научна област: Библиотекарство и информатика

Ужа научна област: Библиотечка информатика, рачунарска лингвистика

УДК број: 026/027:004.62(043.3)

811.163.1'322(043.3)

Metadata Descriptors and Content Descriptors for Information Retrieval in Digital Libraries

Abstract

In this doctoral dissertation we are investigating the performance of information retrieval in digital libraries using various methods – metadata descriptors and content descriptors.

Development of libraries, starting from the traditional, through electronic to digital, has led to the change in methods of information retrieval. Due to the fact that the amount of information has significantly risen over time, data storage types have also changed leading to more analogue library materials being digitized, but also to more born-digital objects. Some characteristics of those objects are significant for information retrieval.

For the purpose of information retrieval, digital objects can be described with the help of metadata descriptors of different types and in different formats depending on the systems in which the digital objects are stored. Likewise, digital objects can also contain content descriptors which implies the usage of advanced technologies for text refinement (optical character recognition, optical segmentation of articles, recognition of named entities), as well as lexical resources and tools that contribute to better and more precise information retrieval in digital libraries.

Analysis of digital object search and information retrieval for those objects has shown that, regardless of all the functionalities offered by different formats for metadata input, as well as digital objects being connected to the Semantic web, metadata based search does not always yield good results, in terms of precision. On the other hand, search over the entire text using metadata descriptors offers incomparably better results, especially in those digital collections where lexical resources of good quality are involved, most notably, morphological dictionaries.

With the goal of improving information retrieval in digital libraries related to Library science and information science, we have enhanced and complemented lexical resources for the Serbian language with Multi-word units (MWUs) – terminological dictionary in the above mentioned area of research, as well as the general dictionary of Serbian, keeping in mind the standards for building electronic terminological dictionaries and terminological data bases. In the process of adding new MWUs to lexical resources, we have analysed texts in a domain corpus and the effectiveness of the existing finite transducers for extraction of MWUs from the domain corpus was checked, after which triple evaluation of the extracted MWUs was performed. Practical work resulted in adding about 600 new MWUs to the terminological

dictionary, and general dictionary of MWUs for Serbian was also enhanced with new units. These complemented dictionaries gave very good results in term extraction tasks.

Based on the conducted research and analyses, we have concluded that advanced methods do aid in more substantial information retrieval. That was also confirmed by comparing the efficiency and precision of information retrieval using advanced methods, as opposed to using common methods. Likewise, we can say that information retrieval aided by content descriptors gives significantly better results compared with the results of metadata descriptor based information retrieval. It is necessary to stress the pertinence of using lexical resources, among which a special place belongs to the morphological dictionaries, which are of utmost importance for languages with rich flexion such is the Serbian language.

Key words: Information retrieval, Digital libraries, Metadata descriptors, Content descriptors, Lexical resources, Electronic dictionaries, Terminological dictionaries, Serbian language

Scientific field: Library science and Information science

Scientific area: Library informatics, Computational linguistics

UDC number: 026/027:004.62(043.3)

811.163.1'322(043.3)

Садржај

Резиме.....	V
Abstract.....	VII
Садржај.....	IX
1. Увод.....	1
1.1 Предмет рада.....	1
1.2 Циљ истраживања	1
1.3 Хипотезе и очекивани резултати	2
1.4 Корпус.....	2
2 Од традиционалне до дигиталне библиотеке.....	4
2.1 Традиционална библиотека	4
2.1.1 Електронска библиотека	5
2.1.2 Електронска библиотека у нашој средини.....	8
2.2 Проналажење информација у традиционалној библиотеци	12
2.3 Пораст информација и промена носилаца информација	17
3 Дигитална библиотека	23
3.1 Дефиниције	23
3.2 Дигитални објекти и њихове особине	28
3.2.1 Идентификатори дигиталних објеката	35
3.2.2 DOI број – карактеристике	37
4 Проналажење информација у дигиталним библиотекама.....	40
4.1 Дескриптори	40
4.2 Дескриптори метаподатака	41
4.2.1 Метаподаци – дефиниције	41
4.2.2 Врсте метаподатака	43
4.2.3 Формати за унос метаподатака	47
4.2.4 Системи за унос података о објектима	58
4.3 Дескриптори садржаја.....	100
4.3.1 Ручно и аутоматско индексирање – разлике	100
4.3.2 Одзив, прецизност и фреквенција појављивања термина	102
4.3.3 Технологије за пречишћавање текста.....	105
4.4 Лексички ресурси и алати.....	112
5 Проналажење информација – анализа	144
5.1 Примери претраге записа у каталозима применом дескриптора метаподатака	144
5.1.1 Претрага записа у каталогу COBIB.SR	144
5.1.2 Претрага записа у каталогу TEL – Европска библиотека	151
5.2 Примери претраге записа у репозиторијумима применом дескриптора метаподатака и дескриптора садржаја	157
5.2.1 Претрага записа у репозиторијуму PHAIDRA.....	157
5.2.2 Претрага записа у Дигиталној библиотеци Филолошког факултета	163

5.3	Примери претраге записа у дигиталним библиотекама применом дескриптора метаподатака и дескриптора садржаја	167
5.3.1	Претрага записа на порталу Еуропеана – колекције домаћих библиотека	168
5.3.2	Претрага записа у дигиталној библиотеци Библиша	182
6	Надградња и допуна лексичких ресурса српског језика	187
6.1	Терминологија	187
6.1.1	Историјски осврт	187
6.1.2	Дефиниција појма	189
6.1.3	Терминологија као наука	190
6.1.4	Стандардизација у области терминологије	193
6.2	Терминологија из области библиотекарства и информатике	197
6.2.1	Допуна терминолошког речника српског језика из библиотекарства и информатике	206
6.2.2	Примена морфолошких речника на корпусу текстова	213
6.2.3	Анализа резултата у допуни терминолошког и општег речника српског језика	223
6.2.4	Анализа примене лексичких ресурса на новом тексту	236
7	Закључак и план за даљи рад	241
7.1	Закључак	241
7.2	План за даљи рад	244
8	Библиографија	246
8.1	Коришћена литература	246
8.2	Рефрентне веб локације	258
9.	Додаци	265
9.1	Списак слика	265
9.2	Списак табела	267
Прилози	268	
Прилог 1 - Вишечлане речи одабране за допуну терминолошког речника из библиотекарства и информатике	269	
Прилог 2а – Вишечлане леме које су се већ налазиле у речнику DELAC	282	
Прилог 2б – Вишечлане леме екстраховане у множини (облик у једнини прихватљив за терминолошки речник)	284	
Прилог 2в – Вишечлане леме екстраховане и у множини и у једнини	286	
Прилог 2г – Вишечлане речи и њихови синоними одабрани за терминолошки речник	287	
Прилог 3а –Вишечлане речи одабране за општи речник српског језика	288	
Прилог 3б – Вишечлане речи екстраховане у множини (облик у једнини прихватљив за општи речник).....	291	
Прилог 4 – Вишечлане речи које су одабране и за терминолошки и за општи речник	292	
Прилог 5а – Избор конкорданци из дисертације I	295	
Прилог 5б – Избор конкорданци из дисертације II	298	
Биографија аутора	301	
Изјаве о докторској дисертацији	302	

1. Увод

1.1 Предмет рада

Предмет рада ове докторске дисертације је истраживање метода проналажења информација у дигиталним библиотекама и могућности унапређења квалитета тог проналажења, односно што већег одзива и боље прецизности постављеног упита за претраживање. Да би се испитале различите методе проналажења информација у дигиталним библиотекама, неопходно је поћи од основних, класичних метода проналажења информација у традиционалној библиотеци и испитати њихову ефикасност.

У оквиру рада, описале се како традиционалне, тако и савремене методе, а посебан акценат биће на дескрипторима метаподатака и дескрипторима садржаја као потенцијалним методама бољег и ефикаснијег проналажења информација. У ту сврху, сваки од поменутих дескриптора биће најпре детаљно описан, а затим ће се описани дескриптори, као и напредне методе (технологије и алати) који прате проналажење информација у дигиталним библиотекама, применити на дефинисаном корпусу и описале се резултати који су добијени претраживањем овим методама.

Фокус рада је на проналажењу информација управо из тог разлога што дигиталне библиотеке треба да понуде боље и шире могућности претраге – преко комплетног текста докумената који су у њима садржани, а и преко метаподатака којима су документи описани. Дигитална библиотека треба да буде много више од колекције докумената која је у дигиталном облику постављена на мрежу, а предмет овог рада јесте и да утврдимо да ли анализиране дигиталне библиотеке одговарају таквим захтевима.

1.2 Циљ истраживања

Циљеви рада на овој докторској дисертацији су:

- утврђивање да ли и како напредне методе помажу у квалитетнијем проналажењу информација;
- упоређивање проналажења информација напредним методама са резултатима проналажења информација применом класичних метода у традиционалним библиотекама.

У сврху остварења основног циља дисертације, користиле се методе и технологије у виду дескриптора метаподатака и дескриптора садржаја за проналажење информација у дигиталним библиотекама.

1.3 Хипотезе и очекивани резултати

У овом истраживању, полазимо од следеће хипотезе – коришћење напредних метода и технологија на корпусу текстова у дигиталним библиотекама омогућава боље и прецизније проналажење информација у односу на проналажење информација у традиционалним библиотекама или дигиталним библиотечким колекцијама на које нису примењене наведене методе и технологије.

Након спроведеног истраживања, очекујемо следеће научне резултате:

- истраживања спроведена у дисертацији и закључци које добијемо могу се користити у смислу побољшања резултата претраживања у другим библиотекама, дигиталним библиотекама, институционалним репозиторијумима, али и дигиталним архивским и музејским колекцијама;
- истраживања могу дати модел побољшаног проналажења информација и успешнијег претраживања садржаја дигиталних колекција које садрже најразличитије врсте докумената;
- истраживања ће допринети и извесном богаћењу општег електронског речника српског језика на основу резултата претраживања дигиталних колекција, а тиме и повећање видљивости српског језика на оба писма у дигиталном окружењу;
- истраживања ће омогућити и изградњу и допуну електронског термилошког речника српског језика из домена библиотекарства и информатике.

1.4 Корпус

Корпус дигиталних објеката и записа на којем ће се вршити истраживања и анализирати резултати претраживања и екстракције информација подразумева:

- каталожке записе у библиографско-каталожкој бази података COBIB.SR;
- каталожке записе електронског каталога „Европске библиотеке“;
- записе и дигиталне објекте у Дигиталном репозиторијуму Универзитета у Београду – PHAIDRA;
- дигиталне објекте у Дигиталној библиотеци Филолошког факултета у Београду;
- дигитализоване документе и мултимедијалне садржаје на порталу Европеана у оквиру пројеката „Библиотеке Европеане“, „Европске колекције 1914-1918“ и „Европске новине“;

- поравнате текстове на српском и енглеском језику у часопису „Инфотека“;
- дигитализоване публикације из домена библиотекарства и информатике аутора проф. др Александре Вранеш¹ и проф. др Цветане Крстев² који су саставни део Корпуса савременог српског језика.³⁴

¹ Публикације: „Високошколске библиотеке“, „Основи библиографије“, „Српска библиографија у области књижевности, језика и библиотекарства у листовима, новинама и часописима од Орфелина до 1941. године : докторска дисертација“, „Од рукописа до библиотеке : појмовник“

² Материјали за бројне предмете из области рачунарства и информатике на Катедри за библиотекарство и информатику Филолошког факултета у Београду, доступно на:
<http://poincare.matf.bg.ac.rs/~cvetana/Nastava/1314/nastava1314.html>

³ Miloš Utvić. „Izgradnja referentnog korpusa savremenog srpskog jezika” (Doktorska disertacija, Filološki fakultet, 2013).

⁴ Cvetana Krstev & Duško Vitas. „Corpus and Lexicon — Mutual Incompleteness.” In: P. Danielsson & M. Wagenmakers (eds), *Proceedings of the Corpus Linguistics Conference*, 14–17 July 2005, Birmingham, , <http://www.birmingham.ac.uk/research/activity/corpus/publications/conference-archives/2005-conf-e-journal.aspx> (приступљено 12.10.2015)

2 Од традиционалне до дигиталне библиотеке

2.1 Традиционална библиотека

Уобичајена, класична дефиниција традиционалне библиотеке каже да она „прикупља, обрађује, чува у својим фондовима и даје на коришћење грађу и информације без обзира на верске, полне, идеолошке, старосне, професионалне и друге различитости, а у циљу ширења образовања, васпитања, културолошког и научног развоја.”⁵ Једна слична дефиниција традиционалне библиотеке каже да је „она јавна или приватна збирка књига; просторија или зграда где се оне чувају; колекција садржински сличних радова, филмова, рачунарских записа, итд. или место где се ти материјали чувају.”⁶ Појам библиотечка грађа „обухвата штампани и на други начин умножени материјал који је својина библиотеке [...] Библиотечку грађу могу чинити књиге, часописи, новине, музичка дела, репродукције уметничких слика и цртежа, картографске публикације, планови насеља, фонодокумента, репрографска грађа, плакати, огласи, фотодокумента, видео и филмски записи, електронске књиге и периодика, базе података, рукописи, меморисана грађа и друге врсте докумената.”⁷

Питер Брофи, описујући пет модела библиотека, као први тип наводи традиционалну библиотеку. Према Брофију⁸, традиционална библиотека има стварну зграду са архитектонским и унутрашњим дизајном; књижни фонд који је тако организован да изађе у сусрет потребама корисника; часописе и новине који омогућавају корисницима да буду у току са областима интересовања; међубиблиотечку позајмицу која повећава богатство библиотеке; рефералне службе које одговарају на потребе и питања корисника; љубазно и образовано особље и атмосферу добродошлице. Такође истиче да и убудуће треба имати у виду предности оваквог модела библиотеке јер ће поједини корисници увек имати потребу за физичким простором библиотеке и физичким носиоцима информација.

⁵Александра Вранеш и Љиљана Марковић. *Од рукописа до библиотеке*. Библиотека (Београд: Филолошки факултет, 2008), 76.

⁶Richard E. Allen (ed.). *The Pocket Oxford Dictionary of Current English*. (Oxford: Clarendon Press, 1984), 421.

⁷Александра Вранеш и Љиљана Марковић. *Од рукописа до библиотеке*. Библиотечка грађа (Београд: Филолошки факултет, 2008), 86-87.

⁸Питер Брофи. „Мерење учинка за библиотеке 21. века.” *Гласник Народне библиотеке Србије* 6,1(2004):210-211.

Нису само потребе појединих корисника за физичким простором и физичким носиоцима информација једини разлози због којих треба имати у виду добре карактеристике традиционалне библиотеке (које наводи и Брофи). Традиционална библиотека има дугу историју постојања, кроз векове се унапредила, унутрашња организација, стандарди који се примењују у пословању и услуге које се пружају корисницима задовољавале су и најразноврсније захтеве и потребе.

Мајкл Горман сматра да традиционална библиотека ”бира, сакупља, и даје приступ свим облицима записаног знања и информација који су важни за њену мисију и заједницу којој служи, и помаже и подучава у употреби тих извора.”⁹ Кад говори о изворима, Горман истиче да традиционална библиотека увек нуди и нове носиоце информација укључујући и електронске изворе. Он оштро критикује позицију да је традиционална библиотека искључиво физичко место које је везано за прошлост и које се ослања на књигу као главни носилац информација.

Ипак, и поред тога што је традиционална библиотека прилагодљива у погледу увођења нових носилаца информација, у овој библиотеци акценат је углавном на коришћењу штампане грађе, најчешће књига (монографских публикација) и новина и часописа (серијских публикација).

2.1.1 Електронска библиотека

Појава првог електронског записа са описом публикације, а касније и локалних и узајамних електронских каталога довела је до увођење електронске библиотеке^{10,11} у традиционалну библиотеку, а самим тим и промену начина рада, приступа информацијама и организације колекција. Значајан помак у еволуцији традиционалне библиотеке начинили су електронски извори као носиоци информација у библиотечкој колекцији.

Међутим, чак и применом нових рачунарских технологија, традиционална библиотека као организована и функционална целина често и даље своје пословање води на устаљен начин и уходаним процедурама које су донекле олакшане применом аутоматизованих процеса. Уместо лисних каталога и каталога акцесија, односно

⁹Мајкл Горман. *Naše neprolazne vrednosti: bibliotekarstvo u XXI veku* (Београд: Филолошки факултет, 2007), 44.

¹⁰Под електронском библиотеком подразумевамо вођење библиотечког пословања електронски подржаним технологијама при чему се посебно мисли на бележење и преношење информација електронским путем.

¹¹„Електронска библиотека– библиотека чије је пословање потпуно електронски опремљено“— Александра Вранеш и Љиљана Марковић. *Од рукописа до библиотеке*. Електронска библиотека (Београд: Филолошки факултет, 2008), 128.

наруџбина, сада имамо електронске каталоге и електронске листе; уместо класичне поште, комуницира се електронском поштом; међубиблиотечка позајмица се често врши слањем чланака или поглавља у електронском облику путем електронске поште, директно на адресу корисника библиотеке. Потребно је било доста времена да се традиционални библиотечки послови потпуно аутоматизују и да електронски подржане технологије добију пуну примену и промене начин традиционалног библиотечког пословања.

У развијеним библиотечким срединама, какве су нпр. САД и Канада, почетак трансформације традиционалне библиотеке увођењем електронски подржаних процедура везује се за крај 60-их и почетак 70-их година 20. века. Огроман помак у том смислу десио се појављивањем стандарда за библиографски опис публикација у MARC¹² формату¹³ који су развијени у Конгресној библиотеци¹⁴ у Вашингтону. Библиотечко пословање у домену каталогизације и класификације публикација, базирано на појави поменутог формата и стандарда, први је библиотечки поступак који је аутоматизован и који је омогућио креирање база података и електронских каталога библиотека, а самим тим и осавремењавање традиционалних библиотечких процеса. Ти први системи били су једнофункционални – подржавали су само једну аутоматизовану процедуру (нпр. каталогизацију публикација или циркулацију фондова – задуживање, раздуживање, резервисање публикација, упис корисника, вођење евиденције о позајмљеним публикацијама, обавештавање корисника о опоменама и сл.).

Предуслов за аутоматизацију свих библиотечких процедура је појава интегрисаних библиотечких система 70-их година 20. века. Интегрисани библиотечки систем представља скуп интегрисаних апликација које подржавају библиотечке послове, као што су набавка, каталогизација, циркулација, кориснички јавно доступни каталози и њихово техничко функционисање.¹⁵ Минимални интегрисани библиотечки систем чине следеће апликације: каталогизација, циркулација и ОПАС¹⁶ (кориснички јавно доступан каталог), а потпуни интегрисани библиотечки систем укључује још и

¹²MARC– Machine Readable Cataloguing

¹³MARC Standards, доступно на: <http://www.loc.gov/marc/>

¹⁴Library of Congress, доступно на: <https://www.loc.gov/>

¹⁵Joan M. Reitz. ODLIS – Online Dictionary for Library and Information Science, Integrated Library System (ILS), <http://www.abc-clio.com/ODLIS/searchODLIS.aspx> (приступљено 09.11.2015)

¹⁶ОПАС – Online Public Access Catalog

послове набавке, међубиблиотечке позајмице, контролу серијских публикација. Ови библиотечки системи примењиви су како на традиционалну, штампану грађу, тако и на савремене, електронске изворе.

Интегрисани библиотечки системи користе јединствену базу података што је и предуслов за њихово постојање, као и више међусобно повезаних апликација са којима аутоматизују различите библиотечке процесе. Јединствена база података и интегрисани ситем омогућавају да су промене настале у једној апликацији истовремено видљиве у другој (нпр. каталожки подаци унети у апликацију за каталогизацију аутоматски су видљиви у јавно доступном каталогу и омогућавају проналажење информација о публикацији; податак да је књига издата унет у апликацију за циркулацију видљив је у јавно доступном каталогу као информација да је публикација заузета и сл.).

Еволуција рачунарских технологија и интегрисаних библиотечких система довела је до тога да данас имамо савремене системе и библиографске сервисе путем којих је могуће преузимати и размењивати каталожке записе (на националном и међународном нивоу) и прегледати и претраживати каталоге једне или више библиотека путем интернета са било ког места на свету.

Један од таквих великих библиографских сервиса је OCLC¹⁷, основан још 1967. године, који данас нуди различите библиографске и друге услуге (каталогизацију и метаподатке, размену извора и информација, подршку за управљање дигиталним библиотекама и системе за библиотечки менаџмент) за преко 16.000 чланица из више од 100 земаља. Један од највише коришћених система које је креирао OCLC је Светски каталог (WorldCat¹⁸). WorldCat је такође настао 1967. године, а први записи су похрањени 1971. године. Пословна политика WorldCat-а подразумева похрањивање и преузимање записа преко овог комерцијалног сервиса. Данас овај сервис користи преко 72.000 библиотека које записе похрањују на преко 485 језика и дијалеката, а WorldCat садржи преко 347 милиона библиографских записа и преко 2 милијарде података о стању фонда.¹⁹

¹⁷OCLC – Online Computer Library Center, доступно на: <https://www.oclc.org/home.en.html>

¹⁸WorldCat, доступно на: <http://www.oclc.org/worldcat/catalog.en.html>

¹⁹OCLC. WorldCat – a Global Library Resource, <http://www.oclc.org/worldcat/catalog.en.html> (приступљено 09.11.2015)

Када анализирамо ове податке и размотримо ситуацију у традиционалним библиотекама пре појаве електронске библиотеке, запазићемо да је електронска библиотека донела немерљив помак у развоју библиотека, промени библиотечког пословања ка олакшавању и убрзавању процеса рада у њима.

2.1.2 Електронска библиотека у нашој средини

У библиотекама у нашој земљи увођење аутоматизованих процеса у библиотечко пословање започело је крајем 80-их година 20. века и до данас се није потпуно завршило. Прапочетак се везује за 1985. годину када је Народна библиотека Србије почела са радом на Студији „Увођење аутоматске обраде података у библиотечко-информациону делатност Народне библиотеке Србије и мрежу библиотека Србије.”

Ова Студија никад није реализована, али је већ 1987. покренут Пројекат Система научних и технолошких информација Југославије (СНТИЈ). Године 1988. објављена је студија изводљивости СНТИЈ-а у којој је, између осталог прописана, институционална структура система. Предвиђено је да систем чине: библиотеке – специјалне, високошколске, универзитетске и националне, ИНДОК и информационе јединице (службе, центри), реферални центри, специјализовани центри научних и технолошких информација, специјализовани системи научних и техничких информација, архиви и музеји, југословенска библиографска агенција (Југословенски библиографски институт), рачунарски сервиси – хостови, али је прецизирано и увођење узајамне каталогизације у библиотеке Југославије у оквиру централизоване рачунарске мреже.²⁰

Израда библиотечко-информационог система (БИС) као дела СНТИЈ-а и развоја организационих решења и програмске опреме поверена је Институту информацијских знаности (IZUM) из Марибора. Израда БИС-а је имала приоритет у изградњи СНТИЈ-а због „општег значаја библиографско-каталогских података о изворима научних и технолошких информација у Југославији, потребе изградње БИС-а на јединственим основама, у складу са међународним и југословенским стандардима, усмеравања потенцијала носиоца изградње специјализованих система НТИ у правцу аналитичко-синтетичке обраде информација у изградњи специјализованих база података, примене процесирања информација [...], изградње фактографских и full-text база података те

²⁰*Sistem naučnih i tehnoloških informacija Jugoslavije : studija izvodljivosti.* (Beograd: Savez inženjera i tehničara Jugoslavije, 1988), 12-15

експертних система.”²¹ IZUM је као софтверску платформу понудио ATLASS чија надоградања је данашњи COBISS²², а у циљу стварања Југословенске библиографско-каталожке базе података (YUBIB). Ово софтверско решење омогућавало је увођење интегрисаног библиотечког система са модулима за каталогизацију, ОПАС и циркулацију.

Упоредо са развојем овог система постоје и други покушаји аутоматизације библиотечког пословања, а међу њима се издваја аутоматизација Библиотеке града Београда.²³ Аутоматизовано пословање у овој библиотеци уведено је још 1987. године када је као платформа уведен рачунарски програм BIBLIS. Програм је израдио тим програмера „Енергодате” заједно са групом стручњака Библиотеке града Београда. Замишљено је да програм омогући аутоматизацију послова набавке (одабир нових наслова, инвентарисање) и обраде (каталогизација и испис библиографија), потом одвојен сегмент за предметизацију, као и циркулацију фонда (упис корисника, издавање публикација, резервација жељене публикације, регистрација прекорачења рока за враћање и испис опомена за кориснике који нису вратили публикације). Такође, програмери су у другој развојној фази програма предвидели координацију набавке између централне библиотеке и огранака, као и аутоматизацију послова у завичајном фонду.²⁴ Овај програм је 1991. године замењен платформом LIBNET за библиотечко пословање која је израђена у програмском језику Clipper, а састојала се од више модула: каталогизације (што је подразумевало и инвентарисање и класификацију, потребне шифрарнике), циркулације (упис и евиденцију чланова, задуживање и раздуживање публикација и слање опомена, статистике уписаних чланова, штампање извештаја) и ОПАС-а (претраживање фонда намењеног корисницима²⁵). Велики недостатак овог система је што формати за унос података

²¹Тomaž Seljak i Tvrтко Šercar. „Prikaz razvoja Sistema naučnih i tehnoloških informacija Jugoslavije (SNTIJ).” U: *Zbornik referata III jugoslovenskog savetovanja Specijalizovani informacioni sistemi na II naučnom skupu o Sistemu naučnih i tehnoloških informacija*, Dubrovnik, 6-8. jun 1990, str. 9-18. (Beograd: Savez inženjera i tehničara Jugoslavije, 1990):15

²²COBISS – Cooperative Online Bibliographic System and Services (Кооперативни онлајн библиографски систем и сервиси), доступно на: http://www.cobiss.net/platforma_cobiss-SR.htm

²³Библиотека града Београда, доступно на: <http://www.bgb.rs/>

²⁴Митко Манџуков и Јасмина Рацковић. „Примена аутоматске обраде података у Библиотеци града Београда.” *Библиотекар* 40,3/4(1988):15-32

²⁵Јасмина Петровић. „Библиотека града Београда.” *Гласник Народне библиотеке Србије*, 4,1(2002):185-202, https://www.nb.rs/view_file.php?file_id=609 (приступљено 09.12.2015)

нису били у складу са међународним стандардима и што UNIMARC формат није био подржан.

До распада Југославије 1991. године у систему ATLASS (данас COBISS) је учествовало 55 библиотека из свих република. Све велике библиотеке из Србије такође су учествовале у овом систему (Народна библиотека Србије, Библиотека Матице српске, Универзитетска библиотека „Светозар Марковић”, као и Југословенски библиографски институт). Упоредо са распадом Југославије долази до престанка учешћа српских библиотека у заједничком систему узајамне каталогизације.

Период од 1992. до 2003. године у нашој библиотечкој средини одликује се радом у локалним базама библиотека без могућности размене записа међу њима, као и појавом других софтверских решења за аутоматизацију библиотечног пословања.²⁶ Захваљујући средствима из два пројекта – „Виртуелна библиотека Србије”²⁷ и „Изградња кооперативне мреже високошколских библиотека у Србији”²⁸ током 2003. године у нашој земљи поново је успостављен систем узајамне каталогизације. Као почетни задатак, формирана је Виртуелна библиотека Србије²⁹ са седиштем у Народној библиотеци Србије која је као софтверско решење за узајамну каталогизацију прихватила већ познату COBISS платформу. Разлог томе је чињеница да су библиотеке оснивачи Виртуелне библиотеке Србије (Народна библиотека Србије, Библиотека Матице српске, Универзитетска библиотека „Светозар Марковић”) имали искуство у раду на овој платформи, да су у време рада у локалним базама радили на првобитној верзији ове платформе, да је програмски пакет имао сучеља на српском језику, те да су због тога руководства установа сматрала да ће се пренос података и поновно повезивање најлакше успоставити управо на COBISS платформи.³⁰ Виртуелној библиотеци Србије као систему узајамне каталогизације у

²⁶Између осталих BISIS који је покренут 1992. на Природно-математичком факултету Универзитета у Новом Саду као српски програм за каталогизацију кога је две године финансирало Министарство за науку Републике Србије и упоредо са њим BIBLIO (на DOS платформи) и WINISIS (на Windows платформи), оба на бази програма ISIS кога је развио UNESCO. Доста је коришћен и програм Пергам који је од 1995. развијао мали тим стручњака на платформи Cliper.

²⁷Пројекат је финансирао Фонд за отворено друштво

²⁸Темпус пројекат UM_JEP 16059/2001 који је финансирала Европска комисија

²⁹Виртуелна библиотека Србије, доступно на: <http://www.vbs.rs/cobiss/>

³⁰Мишљења смо да је пропуштена могућност да библиотеке у Србији у тренутку поновног успостављања система за узајамну каталогизацију добију боље софтверско решење од поменуте

току 2003. и 2004. године прикључило се и петнаестак факултетских библиотека, да би крајем 2015. године 171 библиотека била у овом систему³¹, а укупан број записа био близу три милиона. У складу са добром праксом светских библиотека, систем COBISS омогућава преузимање записа из каталожке базе WorldCat, а узајамни каталог Виртуелне библиотеке Србије саставни је део овог библиографског сервиса.³²

Међутим, иако је електронска библиотека данас широко заступљена у нашој библиотечкој средини, честа је ситуација у библиотекама да су старе праксе само подржане новим технологијама и да пословање није суштински промењено. Неретко се дешава да се увођењем електронске библиотеке послови дуплирају – вођење инвентарних књига и на традиционалан начин и електронским исписом из базе података, додељивање сигнатура и ручним уписивањем последњег текућег броја у књигу сигнатура, а и аутоматским бројачем у бази података, слањем опомена неодговорним корисницима и класичном и електронском поштом, сравњивање пристиглих наруџби и на штампаним и на електронским листама и сл. Чињеница је да је неповерење према електронској библиотеци чешће код искуснијих и старијих библиотечких радника, а да млађи библиотекарски имају много више поверења у нове технологије и њихове могућности.

Ипак, захваљујући већ дугој историји развоја која покрива безмало пет деценија, очигледно је да електронска библиотека омогућава једноставнији рад и организацију послова у једној библиотеци и чини да библиотека буде део ширег система на националном и међународном нивоу. Електронска библиотека је учинила да се од локалног лисног каталога традиционалне библиотеке пређе пут до светског узајамног електронског каталога и да уски круг корисника који је физички морао да дође у библиотеку прерасте у огроман број корисника којима је информација о фонду многих библиотека путем интернета доступна током целог дана са било ког места.

COBISS платформе, у првом реду Aleph који се користи у великом броју европских и светских библиотека. Такође, сматрамо да би један тако велики и развијен систем као што је Aleph омогућио једноставну конверзију података између стандардизованих формата за унос, као и преузимање записа између библиотека (ExLibris Group. Aleph Integrated Library System, доступно на: <http://www.exlibrisgroup.com/category/Aleph>).

³¹ COBISS.SR. Библиотеке чланице – листа свих библиотека, http://vbshome.vbs.rs/sr/o_cobissu/biblioteke-sc.asp (приступљено 09.11.2015)

³² Значајне српске библиотеке са великим фондовима и великим бројем корисника које су остале ван система COBISS су, између осталих, Библиотека града Београда и Градска библиотека у Новом Саду са својим огранцима. Оне своје пословање обављају у систему BISIS.

Ово поглавље ћемо закључити констатацијом да се у развоју библиотечких услуга отишло још много даље. Ако је електронска библиотека тековина 20. века, онда треба указати на то шта нам је, у смислу развоја библиотека, донео 21. век. Термин Библиотека 2.0 означава „потпуну иновацију библиотечких услуга [...] Ако пратимо ову логику, библиотечко пословање до појаве новог концепта могло би се означити као Библиотека 1.0...”³³ Софронијевић у дефинисању Библиотеке 2.0 цитира Хабиба да „Библиотека 2.0 описује подкуп библиотечких услуга пројектованих тако да испуне потребе корисника које су настале као директна или индиректна последица Веб 2.0”³⁴ што се, између осталог, јасно види на примеру омогућавања корисницима да врше различита обележавања (енг. tagging) у електронском каталогу библиотеке. Активно учешће корисника у креирању и промоцији библиотечких сервиса свакако је иновативна тековина савременог доба.

2.2 Проналажење информација у традиционалној библиотеци

Термин „проналажење информација“ (енг. information retrieval – IR) први пут је употребио амерички научник Калвин Нортруп Муерс (Calvin Northrup Mooers 1919-1994) у својој магистарској тези коју је бранио на Институту за технологију у Масачусетсу (Massachusetts Institute of Technology – MIT) 1948. године.³⁵ Објашњавајући појам цитатом из Муерсовог рада³⁶, Оксфордски речник енглеског језика³⁷ наводи да је „проналажење информација трагање за информацијама коришћењем референсних материјала, посебно оних информација које су ускладиштене у рачунарским системима.”

Иако се проналажење информација везује за рачунарске системе, у библиотекама је и пре појаве рачунара било могуће проналазити информације о жељеним

³³ Адам Софронијевић. „Нова парадигма сарадње у библиотекама“ (Докторска дисертација, Филолошки факултет, 2015), 41

³⁴ Michael Habib. *Toward Academic Library 2.0: Development and Application of a Library 2.0 Methodology*. A Master's Paper for the M.S. in L.S degree, School of Information and Library Science of the University of North Carolina at Chapel Hill. (Chapel Hill, N.C., USA, 2006).

³⁵ Eugene Garfield. „A Tribute to Calvin N. Mooers, a Pioneer of Information Retrieval.” *The Scientist* 11, 6(March 17, 1997):9

³⁶ „The requirements of information retrieval, of finding information whose location or very existence is a priori unknown“ – C. N. Mooers

³⁷ Oxford English Dictionary – OED, online edition

<http://www.oed.com/viewdictionaryentry/Entry/95568#eid417511> (приступљено 02.06.2014)

публикацијама. Развој лисних каталога³⁸, класификационих система³⁹, као и коришћење контролисаних речника⁴⁰ и тезауруса⁴¹ у процесима ручног индексирања и обради и опису библиотечке грађе, омогућило је претрагу према формалним критеријумима (алфабетски, односно ауторски каталог⁴²) и према садржинском аспекту (предметни⁴³ и стручни каталог⁴⁴). Оваква врста претраге има своја ограничења јер унеколико зависи од субјективне процене библиотекара приликом доделе предметних одредница, нарочито у ситуацијама када не постоје контролисани речници. Такође, ограничен је и број формалних критеријума према којима је

³⁸ „Каталози на каталошким листићима формата 7,5x12,5 cm“ – Александра Вранеш и Љиљана Марковић. *Од рукописа до библиотеке*. (Београд: Филолошки факултет, 2008), 12.

³⁹ „Груписање људског знања у одговарајуће класе, обележене нумеричким и алфабетским ознакама, које је резултат рада појединца или институције. Успостављање класификационог система у почетку је било резултат рада филозофа, касније библиотекара.“ – Александра Вранеш и Љиљана Марковић. *Од рукописа до библиотеке*. Класификациони систем (Београд: Филолошки факултет, 2008), 168.

⁴⁰ „Утврђена листа пожељних термина из којих каталогизатор или индексер бира одговарајући приликом доделе предметне одреднице или дескриптора у библиографском запису, да би указао на садржај рада у каталогу библиотеке, индексу или библиографској бази података.“ – Joan M. Reitz. ODLIS – Online Dictionary for Library and Information Science, Controlled vocabulary, <http://www.abc-clio.com/ODLIS/searchODLIS.aspx> (приступљено 16.06.2014) (превод аутора)

⁴¹ „Абецедно организован лексикон појмова који је састављен од термина из једне академске дисциплине или области студија, са логичким и семантичким релацијама између термина, а у виду листе предметних одредница или дескриптора који се користе као преферентни термини у индексирању литературе из одређеног подручја. У проналажењу информација, тезаурус може да се користи у ширем смислу за лоцирање надређених и сличних термина уколико корисник жели да прошири проналажење, или у ужем смислу за проналажење подређених термина са циљем веће прецизности претраге. Добро дизајниран тезаурус такође омогућава индексору да одржи конзистентност у додељивању индексног термина документу.“ – Joan M. Reitz. ODLIS – Online Dictionary for Library and Information Science, Thesaurus, <http://www.abc-clio.com/ODLIS/searchODLIS.aspx> (приступљено 12.10.2015) (превод аутора)

⁴² „Основни каталог сваке библиотеке у коме су каталошке јединице распоређене према презимену аутора (индивидуална одредница), имену корпоративног тела (корпоративна, односно колективна одредница), наслову дела (стварна одредница) или називу политичко-територијалне заједнице (формална одредница), у азбучном или абецедном редоследу“ – Александра Вранеш и Љиљана Марковић. *Од рукописа до библиотеке*. Алфабетски каталог (Београд: Филолошки факултет, 2008), 11-12.

⁴³ „Предметни каталог представља попис библиотечког фонда по одређеним појмовима (предметима) израженим одредницама. Према томе, предметни каталог непосредно обавештава коју литературу о одређеним предметима поседује библиотека...” – Дивна Глумац. *Предметни каталог у библиотеци општег типа*. (Београд: Народна библиотека Србије, 1980), 7.

⁴⁴ „Стручни каталог покушава да дело смести у одговарајућу класификациону групу прихваћеног класификационог система, за који је библиотекар дужан да пружи видно, прегледно и систематично објашњење коришћеног вештачког језика. У већини наших библиотека у употреби је систем Универзалне децималне класификације (УДК)...“ – Александра Вранеш и Љиљана Марковић. *Од рукописа до библиотеке*. (Београд: Филолошки факултет, 2008), 286.

претрага могућа. Код дела која имају до три аутора, претрага се врши према првом аутору (упутнице укључују и другог и трећег аутора), а код дела која имају више од три аутора или су анонимна дела (аутор није утврђен), претрага се врши према наслову дела. Према наслову се претражују и серијске публикације, а зборници радова са конференције према називу конференције. Дела везана за неку политичко-територијалну заједницу, претражују се према називу те заједнице. Ту су исцрпљене могућности проналажења информација у лисним каталозима.

Електронски каталог библиотеке, било локални каталог, било јавно доступан на мрежи (OPAC⁴⁵), настао још почетком 70-их, а у нашој средини крајем 80-их година 20. века, даје неупоредиво боље могућности за претрагу и проналажење информација. Будући да се публикације обрађују у неком стандардном формату за библиографски опис (MARC21⁴⁶, Unimarc⁴⁷, COMARC⁴⁸ и сл.), могућ је унос много већег броја података. Формати су структурирани тако да садрже поља и потпоља у која се уносе кодирани и библиографски подаци.

Кодирани подаци, у свим наведеним форматима за библиографски опис, изражени су једномесним, двомесним или тромесним кодовима и односе се на:

- врсту записа (штампана грађа, рукописна грађа, догађај и сл.);
- врсту публикације (монографија, серијска публикација, чланак, некњижна грађа, аудио, видео снимак, електронски извор и сл.);
- језик публикације;
- језик оригиналног дела;
- језик насловне стране;
- језик сажетка;
- државу издавања;
- податке о илустрацијама у публикацији (опште илустрације, географске карте, портрети, нацрти, навигацијске карте, ноте, факсимили, грбови, родословне табеле и сл.)

⁴⁵OPAC – Online Public Access Catalog

⁴⁶ Machine Readable Catalog – MARC21 – format for bibliographical data, <http://www.loc.gov/marc/bibliographic/ecbdhome.html> (приступљено 15.06.2014)

⁴⁷Universal MARC format, Unimarc bibliographic format, <http://www.ifla.org/publications/unimarc-formats-and-related-documentation> (приступљено 15.06.2014)

⁴⁸ COMARC/B format za bibliografske podatke, http://obrazovanje.vbs.rs/DesktopModules/CobissIZO/IzoManualtextLoader.aspx?manual=1_comarc_b_svn&iid=1 (приступљено 15.06.2014)

- податке о томе који је књижевни облик у питању (роман, драма, кратка проза, поезија, антологија, народна књижевност и сл.);
- податке о којој је врсти рада реч (стручни, прегледни, научни, уџбеник, приручник, водич и сл.);
- податке о врсти ауторства (аутор, уредник, преводилац, илустратор, писац предговора или поговора, уметник, фотограф, композитор, прерађивач, рецензент, коментатор, писац, анотација, ментор, члан комисије за одбрану и сл.) итд.

Библиографски подаци су подаци који, по свом садржају, одговарају подацима који су се уносили и на листиће лисног каталога и уносе се према правилима за библиографски опис:

- главни стварни наслов и подаци о одговорности (ауторству);
- издање;
- издавање и штампање – место издавања, издавач, година издања, место штампања, назив штампарије;
- назив колекције;
- материјални опис – број страна, податак о илустрацијама, формат тј. висина хрпта, податак о прилогу уз публикацију;
- напомене – о ауторству, примерку, саставним деловима публикације, предговорима и поговорима, тиражу, библиографијама, регистрима, сажецима и сл;
- предметне одреднице;
- Универзална децимална класификација⁴⁹ – УДК;
- Међународни стандардни број књиге – ISBN⁵⁰, Међународни стандардни број серијске публикације – ISSN⁵¹, Међународни стандардни број музикалије – ISMN⁵² итд.

У домаћој каталожкој пракси, библиографски подаци се уносе према Међународном стандардном библиографском опису (International Standard Bibliographic Description – ISBD).⁵³

⁴⁹UDC – Universal Decimal Classification, доступно на: <http://www.udcc.org/index.php/site/page?view=about>

⁵⁰ ISBN – International Standard Book Number, доступно на: <http://www.isbn.org>

⁵¹ISSN – International Serial Data Number, доступно на: <http://www.issn.org/>

⁵²ISMN – International Standard Music Number, доступно на: <http://www.ismn-international.org/>

Након завршетка процеса каталогизације и класификације одређене публикације, подаци о њој се налазе у електронском каталогу библиотеке. Већина ОПАС каталога има могућност једноставног, изборног и командног претраживања. Једноставно претраживање омогућава претрагу према једном критеријуму, а изборно и командно претраживање према више критеријума коришћењем оператора Булове логичке алгебре.⁵⁴ Претрага каталога путем командног претраживања даје велике могућности у смислу комбиновања различитих критеријума за претраживање. У командном претраживању нужно је, поред Булових оператора, коришћење и префикса за претраживање. Критеријуми за претрагу могу бити: аутор (особа), аутор (корпорација), наслов дела, наслов оригинала, место издавања, година издавања, издавач, врста публикације, језик публикације, лично име као предметна одредница, предметне одреднице уопште, слободно обликоване предметне одреднице, код за врсту ауторства, код за врсту садржаја, земља издавања, кључне речи, идентификациони број записа у каталогу и сл. Примера ради, корисник путем командног претраживања лако може доћи до информације које су то монографске публикације– збирке поезије антологијског карактера преведене са немачког на српски језик, издате у Србији, у периоду од 2010. до 2014. године.

Формирањем оваквих најсложенијих упита лако се може доћи до публикације која је кориснику потребна, без обзира да ли је носилац информације књига, серијска публикација, чланак, некњижна грађа, аудио или видео касета, компакт диск.

Међутим, одзив система ће изостати уколико каталогизатор није тачно унео податке, ако је неки податак пропустио, уколико је направио грешку у куцању

⁵³ IFLA. *Međunarodni standardni bibliografski opis (ISBD) – objedinjeno izdanje* (Beograd: Narodna biblioteka Srbije, 2009), http://www.ifla.org/files/assets/cataloguing/isbd/isbd-cons_2007-sr.pdf (приступљено 15.06.2014)

⁵⁴ „Логички систем који је развио енглески математичар Џорџ Бул (1815-1864): Булова алгебра омогућава кориснику да комбинује речи или фразе које представљају значајне појмове приликом претраживања јавно доступних каталога или библиографског података по кључним речима. Три логички команде (понекад се називају "оператори") су доступни у већини софтвера са претрагу:

- Команда ИЛИ се користи да се прошири проналажење укључујући синониме и сродне изразе у упиту;

- Команда И се користи да се смањи број резултата претраге. Сваки пут када се дода још један концепт коришћењем оператора И претраживање постаје прецизније. У неким јавно доступним каталогима базама података, команда И је имплицитна (нема потребе да се куца између термина). У другим базама, кључне речи ће се претраживати као фраза, ако се не раздвоје оператором И;

- Команда НЕ се користи да искључи нежељене записе из резултата претраге.“ – Joan M. Reitz. ODLIS – Online Dictionary for Library and Information Science, Boolean, <http://www.abc-clio.com/ODLIS/searchODLIS.aspx> (приступљено 12.10.2015) (превод аутора)

приликом уноса података и сл. Корисници традиционалних библиотека сами или уз помоћ библиотекара-информатора постављају захтеве за претраживање у јавно доступним електронским каталозима и библиографским базама података. Често се у традиционалним библиотекама организују и курсеви за кориснике на којима им се показује како да направе сложене упите и како да користе информације из јавно доступног каталога.

Иако електронске каталоге могу претражити са удаљеног места, да би користили физичке примерке библиотечког материјала, корисници лично морају да дођу у библиотеку и жељене материјале преузму. Како је, код традиционалних библиотека, публикације могуће претраживати само преко података који су, током процеса каталогизације и класификације, унети у машински читљив формат у коме је публикација описана, али не и претраживањем целокупног текста публикације, то су многе информације изгубљене, нису видљиве и корисници до њих не могу доћи. „Упркос мноштву секундарних и референсних алата који су развијени за приступ традиционалним документима, проналажење жељене информације у њима је прави изазов.“⁵⁵

2.3 Пораст информација и промена носилаца информација

Носилац информација је „физички медијум у коме или на комесе информације (подаци, звук, слике, итд.) чувају, на пример папир и отисак мастила за штампане материјале, односно магнетна трака или оптички или магнетни диск за електронске изворе. За неке категорије материјала, носилац информација се може трајно сместити у заштитно кућиште од другог материјала (пластика, метал, итд) као саставни део јединице – дискете или зип дискови. Исти подаци могу да се чувају на више типова носилаца, на пример, филм на филмској траци, касети и DVD.“⁵⁶⁵⁷

Кроз историјски развој материјала за записивање и чување података, носиоци текстуалних информација били су папирус, глинене плочице, пергамент, папир; потом носиоци сликовних (визуелних) информација цртежи, фотографије, слике, графике, мапе; појавом радија долазимо до аудио информација на грамофонским плочама и

⁵⁵ Cvetana Krstev. „Digitalne biblioteke kao potencijalni lingvistički resurs: stanje u Srbiji.“ In: Voss, Ch. und Golubović, B. (hrsg.) *Srpska lingvistika – Serbische Linguistik*, SLCCEE, Bd. 7, pp. 123-142. (München, Berlin: Otto Sagner, 2010):123.

⁵⁶DVD (Digital versatile disk) – Дигитални вишенаменски диск

⁵⁷Joan M. Reitz. ODLIS – Online Dictionary for Library and Information Science, Information carrier, <http://www.abc-clio.com/ODLIS/searchODLIS.aspx> (приступљено 16.06.2014) (превод аутора)

аудио касетама, а појавом покретних слика до видео записа и филмова на филмским и видео тракама; електронски подаци складиште се, историјски гледано, на магнетним тракама, дискетама, компакт дисковима, DVD, да бисмо данас дошли до тзв. сторица, односно складишта података (од енгл. storage) који подржавају велику количину података. Развојем носилаца информација, односно њиховом еволуцијом, било је могуће лакше и брже преношење података. Ако само упоредимо рукописну са штампаном књигом видећемо колики је цивилизацијски помак донела штампа и у домену преношења информација. Нова револуција у развоју комуникације и преноса информација дошла је са појавом интернета и са брзим протоком података у електронском облику.

Сви претходно наведени носиоци информација, пре појаве интернета, имали су недостатке у смислу ограничене количине података који су се на њима могли чувати. Меморијски капацитети носилаца информација били су прецизно дефинисани. Тачно се зна колико минута филма стаје на филмску траку или колико минута аудио снимка на грамофонску плочу, односно аудио касету. Дискете⁵⁸, компакт дискови – CD⁵⁹, DVD⁶⁰, иако носе податке у електронском облику, такође имају ограничења у погледу количине похрањених података. Појава интернета омогућила је да се огромне количине информација на једноставан начин објављују, похрањују, претражују, проналазе и користе. Једини неопходан услов је да је информација у електронском облику и да постоји интернет веза. Теоријски гледано, нема ограничења у погледу времена и простора, као ни у погледу количине података.

Међутим, озбиљан проблем у свету електронских информација може да представља и застаривање носилаца дигиталних информација. Под застаривањем ових носилаца подразумева се ”појава у којој дигитални ресурс више не може да се прочита, било зато што физички медијум на коме се налази више није у употреби, или није доступан хардвер или софтвер који се на њему извршава.”⁶¹ Примера застаривања физичких медијума има много – од грамофонских плоча које су замењене аудио-касетама, преко видео касета које су заменили CD и DVD, па све до дискета које су

⁵⁸ Дискета од 8 инча – максимални капацитет 1МВ, дискета од 5¼ инча – максимални капацитет 1.2МВ, дискета од 3½ инча – максимални капацитет 1.44МВ

⁵⁹ Капацитет 650, 700 и 800МВ

⁶⁰ Капацитет од 4,7 до 18GB

⁶¹Сандра Дељанин. „Застаривање дигиталних информација.” *Инфотека* 13,1(2012), 48

застареле појавом најпре CD и DVD, а затим и преносиве USB⁶² меморије. Застарелост хардвера огледа се или у великој истрошености компоненти – читача, чврстог диска и сл. или у потпуној замени одређених делова услед напретка технологије – читач дискете замењен CD или DVD читачем. Такође, како се креатори комерцијалног софтвера залажу за све већу продају својих програмских пакета, велики проблем је што претходна верзија једног програма не подржава документа настала у напреднијој верзији, а још већи што нове верзије софтвера не подржавају податке настале у старијим верзијама, тако да су корисници принуђени да стално купују најновије верзије софтвера и да своје податке конвертују у новије формате.

Застарелост носилаца дигиталних информација, хардвера и софтвера са комплетном проблематиком коју носи, свакако је значајна и незаобилазна тема у свету електронских информација. Корисници модерних технологија који су одушевљени могућностима које оне доносе не треба да забораве да свака технологија застарева и треба да се припреме на то не би ли своје податке сачували од пропадања.

Ипак, развоју науке и експоненцијалном порасту научних информација од друге половине 20. века изузетно је погодовао нови начин похрањивања информација. Осим публикувања резултата научних истраживања у штампаним књигама и часописима, развојем технологије омогућено је публикување у електронским књигама и часописима који се издају на компакт дисковима и DVD, али и публикување на интернету (без физичког носиоца). Такође, електронске публикације доступне преко мреже, омогућавају учење и студирање на даљину⁶³, па је то још један значајан аспект развоја ових садржаја.

Све наведено представља и нове захтеве за библиотеке у којима се информације чувају, без обзира на носиоце информација. „Библиотеке које су увек спремно одговарале захтевима сваког времена, прилагођавају се и новим носиоцима информација, те нуде истовремено и штампани документ и виртуелну информацију.”⁶⁴ „Док су штампани материјали лако преносиви, трајни, опипљиви,

⁶²USB – Universal Serial Bus

⁶³„Учење на даљину – Метод наставе и учења дизајниран тако да превазиђе временске и просторне баријеретакто што омогућава студентима да студирају од својих кућа, на удобан начин, користећи материјале доступне електронским путем или поштом.“ – Joan M. Reitz. ODLIS – Online Dictionary for Library and Information Science, Distance learning, <http://www.abc-clio.com/ODLIS/searchODLIS.aspx> (приступљено 16.06.2014) (превод аутора)

⁶⁴ Александра Вранеш. *Високошколске библиотеке* (Београд: Универзитетска библиотека „Светозар Марковић : Tempus projekat, 2004), 11

можемо их поседовати и враћати им се и у њима естетски уживати, а да нам за то нису потребна никаква техничка и телекомуникациона помагала; електронска издања су актуелна, мултимедијална, флуидна, динамичка, отворена за вишеаспектна, интердисциплинарна и интерактивна претраживања, али истовремено и материјално захтевна.⁶⁵ Из описаних карактеристика електронских издања јасно је да су она, као производ савременог тренутка, изузетно значајна за коришћење, проучавање и истраживање.

На чињеницу да су електронске публикације постале нарочито корисне, како у библиотечко-информационој делатности уопште, тако и за кориснике библиотека, указује и то да је ова врста публикација јаснодефинисана и у међународним стандардима и у домаћој библиотечкој законској регулативи. Међународном стандардизацијом даје се валидност новом типу публикација и оне се стављају у равноправан положај са раније доступним штампаним публикацијама. У приручнику за ISBN дефинисано је да „када је публикација доступна у електронској форми (на пример: е-књига, CD-ROM или публикација доступна на интернету) може да добије ISBN ознаку под условом да садржи текст и да је доступна јавности, и да се не очекује да ће се ажурирати и наставити са излагањем.“⁶⁶ Како и у нашој земљи постоји све више електронских издања публикација, тако је и у домаћем Закону о обавезном примерку, ова врста публикација узета у обзир и јасно дефинисана: „електронска публикација је публикација објављена у електронској форми као посебна физичка целина (дискета, CD, DVD и сл.), публикација која је доступна на интернету и публикација припремљена за штампу у формату који одговара међународним стандардима универзалне доступности информација.“⁶⁷ У истом Закону, осим монографске публикације, као посебна врста серијских публикација наведене су и електронске публикације са континуираним ознакама.

Претрагом у узајамној библиографско-каталожској бази података српских библиотека – COBIB.SR долазимо до информације да је број записа за публикације означене као „електронски извор“ у овом тренутку⁶⁸ близу 15.300, од чега је око

⁶⁵ Александра Вранеш и Љиљана Марковић. *Од рукописа до библиотеке*. Електронско издавање (Београд: Филолошки факултет, 2008), 129.

⁶⁶ ISBN агенција. *ISBN приручник: међународно издање* (London: Међународна ISBN агенција, 2012), 18

⁶⁷ Министарство културе и информисања Републике Србије. Закон о обавезном примерку публикација, члан 4, став 3. (Службени гласник Републике Србије, бр. 52/2011)

⁶⁸ Стање на дан 12.10.2015. год.

12.300 записа за електронске публикације на српском језику. Под овим публикацијама махом се подразумевају оне објављене на CD или DVD, које у око 12% случајева имају и доступ преко интернета (око 1.500 публикација). У нашој земљи, код електронских публикација, тенденција је још увек ка томе да постоји физички носилац, само су подаци у електронској форми. Невелики број издања у нашој земљи публикује се искључиво онлајн иако савремени тренутак налаже све већу потребу за таквом врстом публикација. Разлози за то свакако су нерешена законска регулатива у издаваштву, недовољно познавање ауторског и сродних права, као и чињеница да огроман број електронских монографских публикација постоји на енглеском језику и може се бесплатно користити преко различитих сервиса.

Међутим, овакав тренд мора се брзо мењати јер серијске публикације које излазе у електронском облику, нарочито научни часописи постају последњих деценија толико популарни да библиотеке и појединачни корисници отказују претплату на штампана издања. Информација у научним часописима електронским путем је доступна и најудаљенијем кориснику, а многи нови научни часописи никада нису ни излазили у штампаном облику и користе се искључиво као електронске публикације. Многи од њих су потпуно бесплатно доступни.⁶⁹

Значај научних серијских публикација доступних преко интернета препознат је и у нашој средини још 2001. године, па данас корисници библиотека у Србији имају могућност да преко Конзорцијума библиотека Србије за обједињену набавку – КоBSON користе онлајн преко 35.000 страних научних часописа чији су комплетни чланци доступни и преко 90.000 књига. КоBSON су основале најзначајније научне библиотеке у Србији – Народна библиотека Србије, Универзитетска библиотека „Светозар Марковић“, Библиотека Матице српске, Универзитетска библиотека „Никола Тесла“ из Ниша, Универзитетска библиотека из Крагујевца, Библиотека Српске академије наука и уметности, као и струковно удружење Заједница библиотека универзитета у Србији.⁷⁰ Приступ публикацијама финансира Министарство културе и информисања Републике Србије, а публикације доступне преко КоBSON-а су незаобилазан извор информација за научну заједницу Србије.

⁶⁹Gordon Neavill. Emergence of Digital Libraries. *Infotheca* 5,1-2(2004):30

⁷⁰ Конзорцијум библиотека Србије за обједињену набавку – КоBSON, http://www.kobson.nb.rs/kobson/o_konzorcijumu.1.html (приступљено 16.06.2014)

Зашто су научне и друге публикације у електронском облику доступне преко интернета занимљиве корисницима библиотека? Одговор лежи у особинама тих публикација – ове публикације не могу бити заузете, корисници не морају чекати да буду враћене, доступне су у било ком тренутку, могу их изнова користити са било ког удаљеног места.

Ипак, поред свих добрих особина, електронске публикације имају и негативне стране. Да би се користиле комерцијалне базе података које дају доступност целокупног садржаја часописа, потребно је редовно плаћати претплату која није нимало ниска. Уколико се, из било ког разлога, прекине уговор о коришћењу и откаже претплата, долази се у ситуацију да целе колекције нестану и не могу се више користити.

Међутим, када узмемо у обзир да су електронске публикације, објављене било на физичком носиоцу, било на интернету, постале незаоблазне за најшири круг корисника различитог узраста и интересовања, да се те публикације једноставно користе, да су доступне 24 сата са различитих удаљених места, онда је јасно да је развој и коришћење ових публикација био основ и предуслов за појаву дигиталних библиотека као специјалних колекција најразноврснијег садржаја.

3 Дигитална библиотека

3.1 Дефиниције

У претходном поглављу објаснили смо промену носилаца информација која је условила појаву нових типова библиотечких колекција – електронских публикација, мултимедијалних збирки и сл. Међутим, могућност постављања нових носилаца информација на мрежу носи са собом једну значајну промену у еволуцији библиотека – појаву дигиталне библиотеке. У времену експоненцијалног пораста информација и нових носилаца информација, а узевши у обзир проблеме у претраживању и проналажењу информација везане за традиционалне изворе и традиционалне библиотеке, наметнула се потреба за дигиталним библиотекама путем којих ће корисници брже и једноставније доћи до жељених материјала и информација.

Да бисмо утврдили особине, структуру, организацију, значај, коришћење и улогу дигиталних библиотека, потребно је размотрити различите дефиниције овог појма и анализирати које све аспекте дигиталне библиотеке оне покривају.

У разматрању различитих дефиниција, поћи ћемо од мишљења да је дигитална библиотека део традиционалне библиотеке јер често садржи дигитализоване аналогне садржаје. „Дигитална библиотека је део корпуса и сервиса традиционалне библиотеке који су дигитализовани и који су доступни преко мреже са локалног или удаљеног места.“⁷¹ Сложићемо се са делом дефиниције који се односи на дигитализоване садржаје као саставни део сервиса традиционалне библиотеке. Међутим, колико год да је целокупна дефиниција поткрепљена примерима у пракси, дигитална библиотека, осим наведених, дигитализованих садржаја, садржи и оне материјале која су изворно дигитални или који су настали као дигитални (енг. „digital born“).

Филип Бејкер (Philip Barker) у својој дефиницији из 1994. сматра да треба разликовати електронску, дигиталну и виртуелну библиотеку. Електронска библиотека, према његовом мишљењу, фаворизује приступ документима путем рачунара. Друга – дигитална састоји се само од дигитализованих докумената и у њој нема више традиционалних књига, периодике, новина. А само последња – виртуелна обухвата укупност чулног искуства.⁷² На основу развоја дигиталних библиотека, установили смо, да можемо критиковати овакву рану Бејкерову поделу јер „укупност

⁷¹Abderrazak Mkadmi et Imad Saleh. *Bibliothèque numérique et recherche d'informations* (Paris: Lavoisier, 2008), 112.(превод аутора)

⁷²Philip Barker „Electronic libraries — visions of the future.” *The Electronic Library* 12, 4(1994):221-230

чулних искустава“ може пружити и дигитална библиотека која садржи мултимедијалне садржаје, а најчешћи случај управо и јесте да дигитална библиотека, осим текста, садржи различите материјале – од фотографија или илустрација, до аудио и видео записа.

Поделу на електронске, виртуелне и дигиталне библиотеке истиче и Александра Вранеш. Разматрајући ове појмове, она напомиње да „електронске, виртуелне и дигиталне библиотеке носе разлике, али су све оне синоним брзог информисања које не познаје националне, географске и историјске разлике [...] Дигитална библиотека не означава само дигитализоване колекције, које често почивају на регионалном, завичајном, националном принципу, већ процедуре и околности чија трансформација од традиционалних ка технолошки унапређеним омогућава успостављање дигиталних колекција...“⁷³ У овој опширној дефиницији запажамо синтезу поменутих појмова и таква дефиниција у складу је са развојем дигиталних библиотека као носилаца културног наслеђа, али и најактуелнијим резултатима истраживања у разним научним дисциплинама.

Као и у „Библиотекарском терминолошком речнику“ Љиљане Ковачевић и коауторки⁷⁴, разлику између виртуелних и дигиталних библиотека истиче и Џоан М. Риц (Joan M. Reitz) у свом онлајн речнику из библиотекарства и информатике. Виртуелна библиотека је, према њеној дефиницији, „библиотека без зидова“ – колекције не постоје ни на једном традиционалном физичком носиоцу, али су доступне електронским путем (похрањене су на чврстим дисковима сервера). Такође, Риц сматра да је термин „дигитална библиотека“ прикладнији јер „виртуелност“ подразумева нешто што заправо не постоји. Читање са екрана није исто као читање штампаног издања, али је чињеница да је садржај информација исти без обзира на формат.“⁷⁵ Дигитална библиотека је, према Риц, библиотека у којој су ресурси доступни у машински читљивом формату (на супрот штампаном или микрооблику), а

⁷³ Александра Вранеш. *Високошколске библиотеке* (Београд: Универзитетска библиотека „Светозар Марковић : Tempus пројекат, 2004), 257.

⁷⁴ Љиљана Ковачевић, Весна Ињац и Добрила Бегенишић. *Библиотекарски терминолошки речник* (Београд: Народна библиотека Србије, 2004)

⁷⁵ Joan M. Reitz. ODLIS – Online Dictionary for Library and Information Science, Virtual library, <http://www.abc-clio.com/ODLIS/searchODLIS.aspx> (приступљено 16.06.2014) (превод аутора)

користе се путем рачунара. Дигитални садржај може бити локално доступан или му се може даљински приступити преко рачунарских мрежа.⁷⁶

Вилијам Армс (William Y. Arms) сматра да је дигитална библиотека контролисана односно системски организована колекција информација са придруженим сервисима, при чему су информације ускладиштене у дигиталном формату и може им се приступити преко мреже.⁷⁷ Он истиче технички и информациони аспект дигиталне библиотеке подвлачећи битност информација, информационих сервиса, складиштења информација и удаљеног приступа.

Информатички приступ дигиталној библиотеци, говорећи о информационим системима, претраживању и проналажењу информација, даје и Цветана Крстев: „За све оне кориснике који су укључени у процесе проналажења, дигиталне библиотеке представљају проширење таквих система, које је обogaћено мултимедијалном разноврсношћу и удруженим системима (енг. federated systems). У сваком случају, дигиталне библиотеке морају да подржавају велике колекције докумената, како и могућност њиховог претраживања, каталогизирања и индексирања.“⁷⁸ Нагласићемо то да се Крстев посебно осврће на проналажење мултимедијалних докумената и информација у њима, као саставног дела дигиталних библиотека.

Посматрајући потребе и улогу библиотека, свеобухватну дефиницију дигиталне библиотеке као извора различитих података за кориснике библиотека даје Кристин Боргман (Christine Borgman). Према њеном мишљењу „дигиталне библиотеке су скупови електронских извора и пратећих техничких могућности за креирање, претрагу и коришћење информација. У том смислу, оне су продужетак и унапређење система за за складиштење и проналажење информација који манипулишу дигиталним подацима у било ком облику (текст, слика, звук, статичке или динамичке слике) и постоје у дистрибутивним мрежама. Садржај дигиталних библиотека обухвата податке, метаподатке⁷⁹ који описују различите аспекте података (нпр. представљање, креирање, власништво, репродукцију, права коришћења)...“⁸⁰ Оваква

⁷⁶Joan M. Reitz. ODLIS – Online Dictionary for Library and Information Science, Digital library, <http://www.abc-clio.com/ODLIS/searchODLIS.aspx> (приступљено 16.06.2014) (превод аутора)

⁷⁷William Y. Arms. *Digital libraries*. (Cambridge, Massachusetts, London: The MIT Press, 2000)

⁷⁸ Цветана Крстев. „Дигиталне библиотеке – разграничење појмова.” *Инфотека* 3, 1-2(2002):4

⁷⁹Појам ће бити детаљно описан у поглављу 4.2.1

⁸⁰ Christine Borgman. The invisible library. In: Kresh, D. (ed.). *Whole Digital Library Handbook* (Chicago: American Library Association, 2007):4-5 (превод аутора)

дефиниција јасно означава појам дигиталне библиотеке, њену намену и садржај, начин на који су дигитални подаци описани и веома је блиска „класичним“ дефиницијама библиотека.

Неки аутори, попут Гордона Невила⁸¹, заступају мишљење да је дигитална библиотека везана за аутоматизацију библиотечких система и настанак електронских каталога. Ипак, „права суштина дигиталних библиотека је у самим информацијама, а не у метаинформацијама или референсним информацијама које су у дигиталном облику.“⁸² Управо ова функција дигиталне библиотеке – доступност информација које су садржане у дигиталним објектима – најзначајнија је новина и за библиотекарe и за крајње кориснике. Природа материјала похрањених у дигиталној библиотеци у великој мери олакшава проналажење информација, па се не може сматрати да дигиталну библиотеку чине електронски каталози или дигитализовани лисни каталози. То су само помоћни апарати којима долазимо до физичког или електронског примерка публикације, а раније наведене дефиниције управо истичу јасну разлику између физичких примерака у традиционалним библиотекама и дигиталних објеката у дигиталним библиотекама.

Питер Брофи говори о дигиталној библиотеци кроз поређење са традиционалном библиотеком. Ако класична зграда библиотеке треба да има функционалну архитектуру и дизајн, дигитална библиотека треба да има функционалан дизајн и организацију података; ако традиционална библиотека треба да има добро одабрани књижни и остали фонд, исто важи и за дигиталну библиотеку; међубиблиотечка позајмица је у новим околностима тражење и омогућавање веза ка другим изворима и документима на мрежи; рефералне службе замењене су онлајн сервисима преко којих корисник комуницира са библиотекарком путем електронске поште, блога или друштвених мрежа; љубазно и образовано особље је незаобилазно у свакој врсти пословања, подједнако у традиционалној и дигиталној библиотеци. Поред бројних различитости које постоје између традиционалне и дигиталне

⁸¹ Gordon Neavill. "Emergence of digital libraries." *Infotheka* 5,1-2(2004):25-33.

⁸² Cvetana Krstev. „Digitalne biblioteke kao potencijalni lingvistički resurs: stanje u Srbiji.“ In: Voss, Ch. und Golubović, B. (hrsg.) *Srpska lingvistika – Serbische Linguistik*, SLCCEE, Bd. 7, pp.123-142. (München, Berlin: Otto Sagner, 2010):123

библиотеке, остаје заједничка функција – детаљан и прецизан одговор на захтев корисника у што краћем временском року.⁸³

Проучавањем наведених и коментарисаних дефиниција дигиталне библиотеке запажамо да се у литератури издвајају осим дигиталне, и виртуелна и електронска библиотека чије се улоге у бројним аспектима преклапају. Ипак, без обзира на одређене разлике у дефиницијама и различите углове посматрања, запазићемо да се инсистира на томе да се садржајима приступа са удаљеног места, а у виртуелном простору.⁸⁴ То јесте и највећа предност за кориснике јер корисници дигиталне библиотеке, делећи заједнички виртуелни простор, не морају више физички долазити у библиотеку већ жељене информације могу да пронађу путем мреже са било ког места у свету и у било које време. Дигитална библиотека као системски организована колекција информација са придруженим сервисима при чему су информације ускладиштене у дигиталном формату и користе се преко мреже, за разлику од традиционалне библиотеке, доводи библиотеку кориснику.⁸⁵

Поједини аутори инсистирају на улози корисника у стварању и одржавању дигиталне библиотеке. Тако Карло Мегhini (CarloMeghini) и сарадници сматрају да је дигитална библиотека скуп дигиталних објеката и сервиса који заједници корисника омогућавају приступ и поновно коришћење.⁸⁶ У поменутом чланку истиче се да корисник дигиталне библиотеке треба да буде у стању да самостално пронађе и идентификује жељене дигиталне објекте, сам креира дигитални објект, опише га потребним подацима и направи нову верзију модификовањем постојећег објекта.

На крају поглавља ћемо закључити да је дигитална библиотека велика колекција дигиталних докумената који се у виду различитих дигиталних података (текст, слика, звук, видео, анимација) или њихових комбинација (мултимедија) похрањују на мрежи, описују различитим метаподацима и повезују са другим информационим сервисима. Дигиталну библиотеку чине различити дигитални објекти садржински повезани, било да су дигитализовани анологни документи или изворни дигитални документи. Осим

⁸³ Питер Брофи. „Мерење учинка за библиотеке 21. века.” *Гласник Народне библиотеке Србије* 6,1(2004):210-211

⁸⁴ Abderrazak Mkadmi et Imad Saleh. *Bibliothèque numérique et recherche d'informations* (Paris: Lavoisier, 2008):111-112.

⁸⁵ William Y. Arms. *Digital libraries* (Cambridge, Massachusetts, London: The MIT Press, 2000)

⁸⁶ Carlo Meghini, Nicolas Spyrtatos, Jitao Yang. „A data model for digital libraries.” *International Journal on Digital Libraries* 11(2010):41

претраге података о дигиталним објектима, дигиталне библиотеке омогућавају и претрагу и проналажење информација садржаних у њима. Због тога што се колекције дигиталних објеката похрањују на мрежи, крајњи корисник им може приступати и користити их без временског и просторног ограничења, са могућношћу да сам креира нови дигитални објекат или нову верзију постојећег дигиталног објекта. Наведене особине дигиталних библиотека управо представљају основне разлике између традиционалних и дигиталних библиотека.

3.2 Дигитални објекти и њихове особине

Дигитални објекат је основна јединица дигиталне библиотеке. Уколико извршимо поређење са традиционалном библиотеком, дигитални објекат истог је значаја за дигиталну библиотеку као библиотечка јединица за традиционалну библиотеку. Као што је поменуто у претходном поглављу, дигиталне објекте чине или дигитализовани аналогни документи или документи који су настали као дигитални и само у том облику постоје. Дигиталним објектима приступа се искључиво путем рачунара и у највећем броју случајева, они су доступни преко веба.

Због наведеног начина приступа, неопходно је да дигитални објекти буду похрањени у неком одговарајућем формату. Зависно од природе садржаја, а посматрано из техничког угла, дигитални објекти се јављају у различитим форматима:

- *за текстуалне и сликовне садржаје најчешће се користе:*
- Joint Photographic Experts Group (JPEG) – битмап формат који се углавном користи са приказивање фотографија и других слика у дигиталном облику; овај 8-битни формат користи се у процесу дигитализације код сканирања и приказа страница са текстом; погодан је и за архивирање и за размену јер захтева мале меморијске капацитете⁸⁷;
- Tagged Image File Format (TIFF) – битмап формат који је развила фирма Aldus првенствено за коришћење у стоном и електронском издаваштву; када је фирму преузела Adobe⁸⁸ корпорација и формат је прешао у њено власништво; подржава 8-битне и 24-битне слике; ово је веома флексибилан формат, па се због тога често користи за размену датотека између различитих рачунарских

⁸⁷ Joint Photographic Experts Group (JPEG), доступно на: <http://www.jpeg.org/about.html>

⁸⁸ Adobe, доступно на: <https://www.adobe.com/>

система, а веома је популаран и у графичком дизајну, издаваштву, а и код професионалних фотографа и фотографа аматера⁸⁹;

- Graphic Interchange Format (GIF) – битмап формат који се најчешће користи за приказивање графичких и сликовних садржаја са бојама које су индексирани у Hyper Text Markup Language (HTML) документима и један је од графичких формата који се најчешће користе за размену слика на интернету; то је 8-битни формат који користи палету од максимално 256 боја што представља велико ограничење код савремених веб страница⁹⁰;
- Portable Network Graphics (PNG) – битмап формат који се користи за чување и приказивање слика на интернету, али без губитака при компресији; подржава 24-бита за боју и даје бољи приказ у односу на GIF формат⁹¹;
- Portable Document Format (PDF) – векторски формат који је веома прилагодљив и може се користити на многим платформама и апликацијама; производ је Adobe корпорације; веома прецизно приказује и подржава различите фонтове, распоред различитих елемената на страници, као и векторске и битмап слике; овај формат омогућава и претрагу и навигацију преко линкова; формат се континуирано развија, па је данас у њега могуће уградити и мултимедијалне садржаје⁹², као и векторске карте - мапе за стабилне и мобилне уређаје са могућношћу комбиновања текстуалних и географских података, укључивања и искључивања одређених опција, навигације кроз целокупан документ;
- Portable Document Format Archival (PDF/A) – векторски формат представља ISO стандардизовану верзију PDF формата која се користи за архивирање и чување дигиталних докумената⁹³;
- Microsoft Power Point (PPT) – кориснички графички програм за креирање и приказивање мултимедијалних презентација у виду слајдова у који је могуће уградити текст, слику, аудио и видео записе или њихову комбинацију;

⁸⁹ Adobe TIFF Resources Page, доступно на: <http://partners.adobe.com/public/developer/tiff/index.html>

⁹⁰ Graphic Interchange Format (GIF), доступно на: <http://www.w3.org/Graphics/GIF/spec-gif87.txt>

⁹¹ Portable Network Graphics (PNG) Specification (2nd ed.), доступно на: <http://www.w3.org/TR/PNG/>

⁹² PDF Reference (6th ed.), доступно на: http://www.adobe.com/content/dam/Adobe/en/devnet/acrobat/pdfs/pdf_reference_1-7.pdf

⁹³ PDF Association. PDF/A facts, доступно на: <http://www.pdfa.org/2013/02/pdfa-facts/>

производ је фирме Microsoft⁹⁴ и саставни део програмског пакета Microsoft Office⁹⁵;

- за аудио записе у употреби су:

- Waveform Audio File Format (WAV) – основни аудио формат у оперативном систему Windows који су заједно развили Microsoft и IBM⁹⁶; све Windows апликације које су у вези са аудио садржајима подржавају датотеке у WAV формату⁹⁷;
- Windows Media Audio (WMA) – формат који је развио Microsoft да би превазишао недостатке WAV формата; реч је о сажетом аудио формату који се примарно користи за Windows Media Player⁹⁸;
- MPEG (Moving Picture Experts Group) – 1 AudioLayer 3 (MP3) – формат који је најшире заступљен аудио формат на интернету, ISO стандардизован, али и затворен јер је заштићен патентом; развијен је на Fraunhofer Institute⁹⁹; највеће предности овог формата су компатибилност са другим форматима и могућност да корисници сами одаберу степен компресије аудио датотетке¹⁰⁰;

- за видео записе најчешће се срећу:

- Quick Time Movie (MOV) – формат је развила компанија Apple¹⁰¹, компатибилан је и са Macintosh и са Windows платформом; то је широко распрострањен формат за мултимедију, а користи се за похрањивање филмских и других видео записа, као и аудио записа, мада може да похрани и текстуалне садржаје (нпр. титлове у филмовима)¹⁰²;

⁹⁴Microsoft, доступно на: <https://www.microsoft.com/sr-latn-rs/>

⁹⁵Microsoft Power Point (PPT), доступно на: <https://products.office.com/en-us/powerpoint>

⁹⁶IBM, доступно на: <http://www.ibm.com/us-en/>

⁹⁷Microsoft Corporation. WAVE and AVI Codec Registries, доступно на: <http://tools.ietf.org/html/rfc2361>

⁹⁸Microsoft Corporation. Windows Media Center, доступно на: <http://windows.microsoft.com/en-US/windows/products/windows-media-center>

⁹⁹ Fraunhofer Institute, доступно на: <http://www.iis.fraunhofer.de/en.html>

¹⁰⁰ Fraunhofer Institute. The Story of MP3: How MP3 was invented, доступно на: <http://www.mp3-history.com/>

¹⁰¹ Apple, доступно на: <http://www.apple.com/>

¹⁰² QuickTimeMovie (MOV), доступно на: <http://www.apple.com/quicktime/>

- Audio Video Interleave (AVI) – формат за аудио и видео записе који омогућава синхронизовану репродукцију звука и слике; развила га је компанија Microsoft.¹⁰³

Још једна специфичност дигиталног објекта јесте да „није више стабилан физички ентитет у познатом облику, који је одређен већ у процесу издања, као што је случај са традиционалним документом.“¹⁰⁴ Приликом издавања традиционалне књиге, још током прелома текста јасно се зна који ће бити формат и опсег, хоће ли илустрације бити црно-беле или у боји, каква је врста повеза и сл. Међутим, нестабилност дигиталног објекта у физичком смислу указује на могућност промене самог дигиталног објекта, његове форме, величине, садржаја. Уколико је у питању дигитални објекат за неку сликовну грађу (нпр. фотографију) која је изворно црно-бела и оштећена, такав објекат може бити ретуширан и поправљен путем разноврсних рачунаских програма за графику, а на основу нијанси на црно-белој фотографији, такође се путем графичких програма може додати боја и кориснику представити како је приказ са фотографије заиста изгледао. На дигиталној верзији фотографије може бити дописан и неки текст који ближе објашњава садржај, па се узимајући у обзир све поменуте могућности мењају и форма и величина и садржај једног дигиталног објекта.

Дигитални објекти као носиоци информација у дигиталним библиотекама, имају своје специфичности које их јасно разликују од библиотечких јединица у традиционалним библиотекама:

- Садржај

Специфичност се огледа у томе што се дигитални објекти састоје се од дигиталних текстова, слика, аудио и видеоматеријала или комбинације неких или свих наведених садржаја.

- Стандардизација

Потреба за успостављањем стандарда је неопходна јер се дигитални објекти јављају у различитим форматима, као што је раније и наведено. Приликом

¹⁰³Microsoft Corporation. WAVE and AVI Codec Registries, доступно на: <http://tools.ietf.org/html/rfc2361>

¹⁰⁴Ghislaine Chartron, Brigitte Guyot, Thierry Lafouge, Sylvie Lainé-Cruzel, Geneviève Lallich-Boidin, Marie France Peyrelong et Jean-Michel Salaün. „Le document numérique: un objet fédérateur de recherche en sciences de l’information.” *Documentaliste – Sciences de l’information* 39, 6(2002),303,<http://www.adbs.fr/le-document-numerique-un-objet-federateur-de-recherche-en-sciences-de-l-information-15075.htm?RH=REVUE>(приступљено 27.06.2014) (превод аутора)

успостављања стандарда морају се узети у обзир различите структуре података похрањених у дигиталним библиотекама, информатички гледано, а свакако и њихово повезивање са другим базама података, мрежама хипертекстуалних докумената¹⁰⁵ и семантичким вебom¹⁰⁶. Коришћење стандарда је неопходно да би се омогућила и размена података између дигиталних библиотека што се нарочито односи на преузимање одређених дигиталних објеката и података о њима јер су саставни део двеју или више дигиталних библиотека. Преузимање скраћује време потребно за креирање дигиталне библиотеке, а тиме се избегава и поновна дигитализација истоветног аналогног садржаја.

- Технологија

Технолошка разноврсност докумената у дигиталним библиотекама је велика, а та особина дигиталних објеката омогућава и већу дистрибутивност њиховог садржаја. Дигитални објекти су описани у стандардним форматима, за њихову дигитализацију користе се стандардни формати. Честа је ситуација да су формати за опис компатибилни у различитим системима, па то омогућава и размену података међу системима. У оваквим случајевима, дистрибутивност не изостаје јер се дигитални објекти настали коришћењем различитих технологија могу користити у различитим системима и досегнути најшири круг корисника. Такође се на овај начин избегава вишеструка дигитализација истог аналогног објекта који је саставни део неколико дигиталних библиотека.

- Ажурирање

За разлику од традиционалне библиотеке где је библиотечке јединице немогуће мењати, а да се не оштете или да се значајно не утиче на њихов садржај, у дигиталним објектима садржај се може променити на једноставан начин – информације се лако мењају и освежавају у ситуацијама кад постоји потреба за тим (уочене грешке у куцању, нетачно описани документи, нове информације о документима и сл.).

¹⁰⁵ „Метод представљања дигиталних информација који омогућава да се одговарајуће фајлови и елементи података међусобно повежу, а не да се посматрају само линеарно. Текстуалне везе уграђене у документу писаном у HTML омогућавају да се информације претражују нелинеарно, асоцијативним принципом слично начину на који функционише људски ум, избором путем курсора или помоћу компјутерске тастатуре. Хипертекст је основни организациони принцип Светом раширене мреже.“ - Joan M. Reitz. ODLIS – Online Dictionary for Library and Information Science, Hypertext, <http://www.abc-elio.com/ODLIS/searchODLIS.aspx> (приступљено 13.10.2015) (превод аутора)

¹⁰⁶ „Semantic web is an extension of the current web in which information is given well-defined meaning, better enabling computers and people to work in cooperation.” – Tim Berners-Lee, James Hendler, Ora Lassil. „The Semantic Web.” *Scientific American Magazine* (May 17, 2001)

- Вишејезичност

Велики број дигиталних објеката јавља се на два или на више језика – најчешће су материјали на језику земље из које је установа која уређује дигиталну библиотеку, па се код проналажења информација морају узети у обзир све особине конкретног природног језика. Иако је енглески језик доминантан у комуникацији на интернету, „предвиђа се да ће управо развој дигиталних библиотека увећати присуство и других језика на интернету.“¹⁰⁷ Већ данас је случај да су осим енглеског на интернету присутни и бројни други језици, па и они са релативно малим бројем говорника. Према интернет статистици о коришћењу језика на веб странама уочавамо да број веб страна који садрже српски језик заузима 0,1% од укупног броја веб страна који се појављују на бар два језика.¹⁰⁸ Уколико се дигитални објекат појављује на више језика, претраживање и проналажење информација о њему постаје изазовније јер је потребно омогућити превођење упита што се постиже употребом посебних електронских речника¹⁰⁹, тезауруса, али и семантичких мрежа¹¹⁰.

- Доступност

Кад се дигитални објекат једном постави на мрежу, постаје јавно доступан под одређеним условима коришћења¹¹¹, а како дигитална библиотека нема радно време, садржаји су стално доступни без временског и просторног ограничења.

- Мултимедијалност

Садржај дигиталних објеката често је мултимедијалан¹¹² (спој текста, слике, анимације, аудио и видео материјала), па се у дигиталним библиотекама мора обезбедити и претраживање мултимедијалног садржаја. Претраживање

¹⁰⁷ Цветана Крстев. „Дигиталне библиотеке – разграничење појмова.“ *Инфотека*, 3, 1-2(2002):5

¹⁰⁸ „Usage of content languages for websites“, http://w3techs.com/technologies/overview/content_language/all (приступљено 14.10.2015)

¹⁰⁹ Појам „електронски речник“ биће детаљно објашњен у одељку 4.4

¹¹⁰ „A semantic network or net is a graph structure for representing knowledge in patterns of interconnected nodes and arcs. Computer implementations of semantic networks were first developed for artificial intelligence and machine translation, but earlier versions have long been used in philosophy, psychology, and linguistics.“ – John F. Sowa. *Semantic Networks*. In: Stuart C. Shapiro (ed.) *Encyclopedia of Artificial Intelligence* (Hoboken, NJ, Wiley, 1992), <http://www.jfsowa.com/pubs/semnet.pdf> (приступљено 14.10.2015)

¹¹¹ Дефинисано према захтевима креатора дигиталне библиотеке, а уређено према конкретном закону о ауторском праву и интелектуалној својини.

¹¹² „Multimedija sadrži dva ili više audiovizuelna izražaja, na primer zvuk i sliku, tekst i animiranu grafiku.“ – IFLA. *Smernice za audiovizuelnu građu u bibliotekama i drugim institucijama* (Hag : IFLA, 2004), 4, <http://archive.ifla.org/VII/s35/pubs/avm-guidelines04-sr.pdf> (приступљено 27.06.2014)

мултимедијалних садржаја битно је другачије од претраживања текстуалних садржаја. Поређења ради, претраживање слика врши се упитима на нивоу визуелних обележја: боје, текстуре, облика, локације; претраживање по одређеном типу објекта, конкретном објекту или особи; претраживање по апстрактним атрибутима: типу догађаја или активности, емоционалним стањима.¹¹³

Када се говори о особинама дигиталних објеката, значајна је и категоризација наведена у раду Јаниса Калиникоса (Jannis Kallinikos) и сарадника.¹¹⁴ Они издвајају четири особине дигиталних објеката:

1. Уређивање

Дигитални објекти се могу брзо и лако модификовати. Могуће је преуредити елементе од којих је састављен дигитални објекат – неке ставке се могу брисати, мењати или се могу додавати нове. Код редовног ажурирања дигиталне библиотеке уређивање игра значајну улогу јер се на једноставан начин могу уклонити они објекти који се уопште не користе, а додати они за које се сматра да ће бити коришћени. Особина дигиталних објеката да се лако могу уређивати и мењати у потпуном је контрасту са особинама физичких библиотечких или архивских јединица или артефаката у музејским колекцијама – ово је можда и кључна особина која иде у прилог дигиталним објектима у односу на физичке носиоце.

2. Интерактивност

Због своје природе, дигитални објекти нуде различиту врсту коришћења, било целих објеката, било њихових делова, било веза ка другим објектима. Интерактивност се везује за начин коришћења објекта, а не за његову модификацију. Услед могућности интеракције, сваки корисник проналази сопствени начин кретања кроз дигитални садржај, начин на који чита, гледа или слуша информације похрањене у дигиталном објекту. У том смислу, интерактивност дигиталних објеката је у блиској вези са интерактивним коришћењем услуга на вебу јер је кориснику остављена слобода да одабере сопствени правац кретања и проналажења адекватних информација.

3. Отвореност

¹¹³John Eakins and Margaret Graham. *Content-based information retrieval* (Manchester: JISC Technology Application Programm, 1999), 7-8

¹¹⁴Jannis Kallinikos, Aleksii Aaltonen and Atila Marton. „A theory of digital objects.” *First Monday* 15, 6-7, (June 2010), <http://firstmonday.org/ojs/index.php/fm/article/view/3033/2564> (приступљено 27.06.2014)

Отвореност је особина дигиталних објеката која се односи на могућност да их заједница корисника мења. Често се креирају дигиталне библиотеке које корисницима пружају могућност да описују дигиталне објекте, постављају своје дигиталне објекте, врше измене у опису постојећих објеката, мењају дигиталне објекте у смислу побољшавања њиховог садржаја (нпр. постављање фотографије са бољом резолуцијом од оне која се већ налази у дигиталној библиотеци или квалитетнијег звучног или видео снимка). Отвореност и могућност промене од стране корисника дефинисана је унапред при успостављању дигиталне колекције.

4. Дистрибутивност

Интерактивност и отвореност омогућавају бољу дистрибутивност дигиталних објеката. Квалитетни дигитални објекти, тачно описани, постављени у стандардизованом формату и повезани са другим изворима на мрежи омогућавају лако и брзо коришћење и може се сматрати да такви „дигитални објекти немају границе“¹¹⁵ у смислу дистрибутивности њиховог садржаја и приступа до корисника. Ово их јасно одваја од омеђених информација садржаних у физичким носиоцима.

У извесном смислу, категоризација Калиникоса и сарадника садржана је и у описаним специфичностима дигиталних објеката наведених раније (садржај, технологија, ажурирање, доступност). У оба случаја истиче се разноврсност садржаја, могућност његове промене, учешће корисника, боља доступност и дистрибутивност захваљујући технологијама које се користе и специфичним особинама дигиталних у односу на физичке објекте.

3.2.1 Идентификатори дигиталних објеката

У традиционалној библиотеци, библиотечке јединице су означене сигнатурама¹¹⁶ које су јединствени идентификатори путем којих се приступа конкретном физичком објекту (библиотечкој јединици – књизи, часопису, компакт диску, мапи, рукопису, аудио или видео касети и сл.) и омогућава његово проналажење и коришћење.

¹¹⁵Jannis Kallinikos, Aleksii Aaltonen and Atila Marton. „A theory of digital objects.” *First Monday* 15, 6-7, (June 2010), <http://firstmonday.org/ojs/index.php/fm/article/view/3033/2564> (приступљено 27.06.2014)

¹¹⁶ „Сигнатура – Ознака која одређује место публикације на библиотечкој полици. Појам се развио од првобитног означавања технике пагинарања редоследа табака словним или нумеричким ознакама.“ – Александра Вранеш и Љиљана Марковић. *Од рукописа до библиотеке. Сигнатура* (Београд: Филолошки факултет, 2008), 277.

У дигиталним библиотекама „сигнатура“ дигиталног објекта јесте идентификатор дигиталног објекта – DOI¹¹⁷. Настанак DOI система иницирала су три издавачка удружења– Међународна асоцијација издавача¹¹⁸, Међународно удружење научних, техничких и медицинских издавача¹¹⁹ и Удружење америчких издавача.¹²⁰ Иако је првобитно настао у свету издаваштва, DOI је замишљен као генерички оквир за управљање идентификованим садржајима дигиталних мрежа, узимајући у обзир тенденцију развоја дигиталних садржаја и њихове мултимедијалне доступности. Систем је представљен на Сајму књига у Франкфурту 1997. Исте године формирана је Међународна DOI фондација¹²¹ са циљем да развија и управља DOI системом.¹²² DOI систем се наслања на добру праксу других идентификационих система, као што је у првом реду систем Међународног стандардног броја књиге¹²³ и других стандардних бројева из света издаваштва (Међународни стандардни број часописа¹²⁴, Међународни стандардни број за аудио и видео записе¹²⁵).

DOI је име ентитета (објекта) у дигиталном свету на мрежи. Он је битан за идентификацију и размену информација на мрежи, „обезбеђујући тако компатибилност са другим стандардним бројевима и везу са интернет адресом на којој је извор објављен.“¹²⁶ DOI број садржи метаподаци о дигиталном објекту, а један од тих података је и URL¹²⁷ адреса. Битно је нагласити да је DOI број стабилнији од URL адресе. Уколико се због промене сервера или назива домена URL адреса промени, издавач у метаподацима DOI броја треба само да изврши корекцију и садржај је поново доступан.

¹¹⁷DOI – Digital Object Identifier, <http://www.doi.org/> (приступљено 30.06.2014)

¹¹⁸ International Publishers Association

¹¹⁹ International Association of Scientific, Technical and Medical Publishers

¹²⁰ Association of American Publishers

¹²¹International DOI Foundation (IDF)

¹²² International DOI Foundation. „DOI Handbook, History“, https://www.doi.org/doi_handbook/1_Introduction.html (приступљено 14.10.2015)

¹²³International Standard Book Numer – ISBN

¹²⁴International Standard Serial Number – ISSN

¹²⁵International Standard Recording Code – ISRC

¹²⁶Александра Вранеш и Љиљана Марковић. *Од рукописа до библиотеке*. Идентификатор дигиталних садржаја (Београд: Филолошки факултет, 2008), 145.

¹²⁷URL – Uniform Resource Locator

3.2.2 DOI број – карактеристике

Системом DOI бројева управља Међународна DOI фондација у Оксфорду¹²⁸, конзорцијум који је отворен за чланство како комерцијалних тако и некомерцијалних партнера. DOI систем, кога је прихватила Међународна организација за ISO стандардизацију¹²⁹, има неколико значајних компоненти:

- прецизну алфанумеричку синтаксу записивања;

Сваки DOI број је јединствен и идентификује један једини ентитет (најчешће чланак у часопису). DOI број се формира од симбола који се састоје од префикса (регистранта) и суфикса (специфичног објекта).¹³⁰ Нема ограничења у погледу дужине једног DOI броја, али је јако важно да се он формира од карактера кодираних према UTF-8¹³¹ који представља имплементацију Unicode-а.¹³² Илустрације ради, показаћемо пример једног DOI броја у домаћем систему doiSerbia: DOI број је 10.2298/JFI1369009P при чему је 10.2298 префикс односно регистрант, а JFI1369009P суфикс, односно специфични објект.¹³³ За све часописе из Србије префикс је исти – 10.2298, а суфикс се састоји од скраћенице наслова часописа, године издавања часописа, броја свеске у тој години, почетне странице на којој се чланак налази и првог слова презимена аутора.¹³⁴

- моделе уноса метаподатака;

¹²⁸IDF – International DOI Foundation, http://www.doi.org/doi_handbook/7_IDF.html (приступљено 30.06.2014)

¹²⁹ISO – International Standard Organization, <http://www.iso.org/iso/home/standards.htm> (приступљено 30.06.2014)

¹³⁰International DOI Foundation., „DOI – Digital Object Identifier. Numbering“, http://www.doi.org/doi_handbook/2_Numbering.html (приступљено 30.06.2014)

¹³¹Један од три најзначајнија Unicode трансформационих формата у коме се свака 21-битна вредност оредставља као секвенција од једног до четири бајта – Svetana Krstev. „Kurs iz XML-a.“, <http://poincare.matf.bg.ac.rs/~cvetana/kurs-xml/xml-codes/xmlcodes12-2.html> (приступљено 15.10.2015)

¹³²Unicode је систем кодирања који су развили истраживачи из ХеросПарс и Apple-а са циљем да буде универзалан, јединствен и униформан – Svetana Krstev. „Kurs iz XML-a.“, <http://poincare.matf.bg.ac.rs/~cvetana/kurs-xml/xml-codes/xmlcodes8.html> (приступљено 15.10.2015)

¹³³Префикс 10.2298 односи се на часописе из Србије, а суфикс JFI1369009P на одређени чланак у оквиру конкретног часописа при чему се суфикс разлаже на следећи начин: JFI –Јужнословенски филолог (скраћеница наслова часописа), 13 – година издавања часописа је 2013, 69 – број свеске је 69, 009 – број почетне стране чланка је 9, P – прво слово презимена аутора (у овом случају Пипер)

¹³⁴Biljana Kosanović. DOI u Srbiji, <http://www.doiserbia.nb.rs/img/DOI2010.pdf> (приступљено 15.10.2014)

DOI као идентификатор нема корисност без метаподатака који описују оно што се идентификује. Минимални сет метаподатака потребних за DOI број дефинисан је као DOI Kernel метаподаци чија је основна функција да омогуће с једне стране препознавање конкретног објекта, а са друге стране размену података међу системима.¹³⁵

- процедуре за имплементацију DOI бројева.¹³⁶

Ове процедуре подразумевају кораке који су неопходни да би једна организација постала чланица DOI система и имала могућност да за потребе својих издања додељује DOI бројеве. Процедура је, између осталог, садржана у следећем: додељивање DOI префикса, покретање и коришћење локалног сервиса за руковање DOI подацима, регистравање DOI имена са припадајућим метаподацима, пренос DOI имена од једног до другог регистранта, препоруке о условима које треба да испуни локални сервис.¹³⁷

DOI бројеви као идентификатори дигиталних објеката, изузетно су значајни у свету научних информација и научних чланака. DOI број је углавном саставни део сваког научног чланка похрањеног на мрежи у оквиру неке дигиталне библиотеке научних радова или научних часописа. И у нашој научној продукцији и код издавача научних часописа запажа се тенденција доделе DOI бројева дигиталним објектима. У свету целокупан посао на пољу DOI бројева раде издавачи научних часописа који потписују уговоре са сервисима за доставу и размену података какав је рецимо CrossRef.¹³⁸ Преко CrossRef-а даље се врши размена и повезивање чланака опремљених DOI бројем са другим референтним системима. Због тога што је у нашој земљи уређивачки рад на научном часопису најчешће волонтерски, а посао око креирања, депоновања, ажурирања и одржавања метаподатака обиман и сложен, додела DOI бројева у Србији обавља се под покровитељством Народне библиотеке Србије и КоБСОН-а.¹³⁹ Међутим, Народна библиотека Србије не мора бити једино

¹³⁵International DOI Foundation. „DOI – Digital Object Identifier. Operating procedures“, http://www.doi.org/doi_handbook/9_OperatingProcedures.html (приступљено 15.10.2015)

¹³⁶Norman Paskin. Digital Object Identifier for scientific data. *Data Science Journal* 4, 1(18 March 2005):1, <http://www.doi.org/topics/041110CODATAarticleDOI.pdf> (приступљено 30.06.2014)

¹³⁷International DOI Foundation. „DOI – Digital Object Identifier. Data model“, https://www.doi.org/doi_handbook/4_Data_Model.html (приступљено 15.10.2015)

¹³⁸CrossRef, доступно на: <http://www.crossref.org/>

¹³⁹ Biljana Kosanović. „Научне информације у Србији – понуда искуства КОБСОН-а.“ *Kultura* 129(2010):112, <http://scindeks-clanci.ceon.rs/data/pdf/0023-5164/2010/0023-51641029106K.pdf> (приступљено 30.06.2014)

место за доделу DOI броја у нашој земљи. Свака установа која потпише уговор са асоцијацијом CrossRef и прихвати услове које прописују CrossRef и Међународна DOI фондација, може да додељује DOI бројеве својим издањима. Посао те врсте започео је Филолошки факултет у Београду за своја бројна научна издања.

Због наведених и описаних особина дигиталних библиотека, дигиталних објеката и информација садржаних у њима, као и идентификатора дигиталних објеката, јасно је да се и проналажење информација у дигиталним библиотекама значајно разликује од проналажења информација у традиционалним библиотекама. Дигитални објекти се могу претраживати на исти начин као и аналогни извори у традиционалним библиотекама – путем података у каталошким записима јавно доступних електронских каталога, али природа дигиталних извора омогућава и проналажење информација у целокупном тексту, као и преко слика и других мултимедијалних садржаја. Због природе дигиталних објеката, њих је могуће „опремити“ најразличитијим подацима, али и на њима применити различите методе и технологије за побољшано и квалитетније проналажење информација.

Главна хипотеза ове докторске дисертације јесте да се у корпусу текстова у дигиталним библиотекама коришћењем напредних метода и технологија омогућава боље и прецизније проналажење информација у односу на проналажење информација у традиционалним библиотекама или дигиталним библиотечким колекцијама на које нису примењене наведене методе и технологије. Због тога ће све ове специфичности, методе и технологије, бити најпре детаљно објашњене у следећем поглављу, да би се у каснијим поглављима њиховом применом на узорку садржаја различитих дигиталних библиотека и потврдила постављена хипотеза.

4 Проналажење информација у дигиталним библиотекама

4.1 Дескриптори

Да бисмо уопште могли да претражујемо и проналазимо информације у базама података, било о аналогним, било о дигиталним документима, неопходно је да они буду описани и означени на неки начин. У том смислу, незаобилазна је улога дескриптора како у традиционалним, тако и у дигиталним библиотекама. „Дескриптор је реч или фраза (индексни термин) који се користи да идентификује јединицу (као субјекат или документ) у системима за проналажење информација.“¹⁴⁰ Дескриптор је, такође, једнозначно одређена реч или група речи, односно термин природног језика који се користи за индексирање докумената и њихову класификацију и претраживање. У домену индексирања, дескриптор јесте најчешће коришћен термин, ознака или низ симбола који је придружен одређеном документу и представља приступну тачку у процесу проналажења информација.¹⁴¹

У традиционалним библиотекама, каталожке базе садрже податке о публикацијама које одређена библиотека поседује. За опис тих публикација користе се дескриптори. Дигиталне библиотеке садрже податке о документима који су у њима похрањени. За њихов опис такође се користе дескриптори. Дескриптори, као идентификатори аналогних објеката у традиционалним библиотекама и дигиталних објеката похрањених у дигиталним библиотекама, одређују формалне и садржинске особине објекта и деле се на:

- дескрипторе метаподатака или објективне дескрипторе и
- дескрипторе садржаја или субјективне дескрипторе.

Дескриптори метаподатака су, између осталих, име аутора, наслов документа, назив издавача, година издавања, место издавања, односно нека формална особина документа.¹⁴² Они се могу применити на различитим носиоцима информација и на различитим форматима информационог извора. Додела метаподатака објектима у дигиталним библиотекама веома је слична каталогизацији у традиционалним библиотекама. Начин додељивања дескриптора метаподатака регулисан је различитим

¹⁴⁰Merriam-Webster Dictionary. Descriptor, <http://www.merriam-webster.com/dictionary/descriptor> (приступљено 7.7.2014)

¹⁴¹Joan M. Reitz. ODLIS – Online Dictionary for Library and Information Science. Descriptor, <http://www.abc-clio.com/ODLIS/searchODLIS.aspx> (приступљено 07.07.2014)

¹⁴² Појам „метаподатак“, његова улога и врсте, биће објашњени у наредном поглављу

усвојеним стандардима и каталошким правилима. У српској пракси, користе се међународни стандарди, тако да је размена података са другим системима олакшана.

Дескриптори садржаја односе се на садржај информације или текста. За разлику од дескриптора метаподатака, код дескриптора садржаја не постоје јасно дефинисани стандарди ни у иностраној ни у домаћој пракси. Код доделе дескриптора садржаја нужно је вршити индексирање документа, а оно може бити ручно или аутоматско. У поглављу 4.3.1 описаћемо и дефинисаћемо и ручно и аутоматско индексирање, али ћемо анализе текстова и претраживања вршити путем аутоматског индексирања. Због квалитетније доделе дескриптора садржаја и проналажења информација, анализираћемо технологије и алате за пречишћавање текста које омогућавају бољи одзив и већу прецизност добијених резултата након постављеног упита, као и на лексичким ресурсима и алатима који су неопходни да би подржали овакву врсту претраживања и проналажења и екстракције информација.

4.2 Дескриптори метаподатака

4.2.1 Метаподаци – дефиниције

Термин метаподаци на различите начине се користи за разноврсне потребе. У класичном библиотекарству, термин метаподаци користи се за формалну структуру описа извора, примењену и на дигиталне и на аналогне документе. Каталогизација у библиотекама представља један облик доделе метаподатака, а формати UNIMARC¹⁴³ и MARC21¹⁴⁴ са правилима који их прате (ISBD¹⁴⁵, AACR2¹⁴⁶) су стандарди за метаподатке. Ови стандарди за метаподатке били су основа за развој каснијих схема и других стандарда који се баве искључиво дигиталним документима. У случају дигиталне библиотеке, реч је о дигиталном извору (дигитализованом аналогном или изворном дигиталном објекту).

Метаподаци су, према Гејл Хоџ (Gail Hodge)¹⁴⁷, структуриране информације које описују, објашњавају, идентификују, лоцирају или на други начин чине лакшим проналажење, коришћење или управљање неким извором информација. У најширој

¹⁴³UNIMARC – Universal MARC format – Универзални МАРК формат

¹⁴⁴MARC21 – Machine-Readable Cataloging 21 – Машински читљива каталогизација 21

¹⁴⁵ISBD – International Standard for Bibliographic Description – Међународни стандард за библиографски опис

¹⁴⁶AACR2 – Anglo-American Cataloguing Rules 2 – Англо-америчка правила за каталогизацију 2.

¹⁴⁷Гејл Хоџ. „Метаподаци на лак начин“. *Гласник Народне библиотеке Србије* 1(2004):157

дефиницији, метаподаци су подаци о подацима или информације о информацијама и углавном се односе на дигитално окружење. Изузетно је битно поново истаћи да су то структуриране информације, јер се због размене података мора користити нека унапред утврђена структура. И Тим Бернерс-Ли (Tim Berners-Lee) сматра да је метаподатак машински читљива информација о веб или другим ресурсима са јасно дефинисаном семантиком и структуром.¹⁴⁸ Слободно обликован, нестандардизован опис ни у класичној каталогизацији ни у домену метаподатака не може се користити за квалитетну размену информација.

Вранеш и Марковић сматрају да је метаподатак „машински читљив податак који се односи на друге податке, на информације о изворима и њиховим ауторима, на библиографске и каталожке информације.“¹⁴⁹ Овим се јасно дефинише да је метаподатак каталожки податак о физичком или дигиталном објекту и да је у том смислу близак каталогизацији. Јако је важно истаћи ову особину метаподатака пошто постоје мишљења да су дескриптори метаподатака дигиталних објеката различити од дескриптора метаподатака аналогног објекта јер је каталогизација нешто што припада физичком објекту, а метаподатак је искључиво везан за електронски извор.¹⁵⁰

Да је додела метаподатака блиска каталогизацији помиње и Крстев тврдећи да се у пракси „метаподаци најчешће користе за опис дигиталних објеката па имају сличну улогу као и записи библиотечких каталога.“¹⁵¹

Дефинишући појам метаподатак, Крстев и Витас истичу да метаподаци користе изабрани формат и коришћењем изабраног формата представљају оно што је њиме забележено те одређују шта се од података може препознати, а „стандардни формат и стандардни мета-подаци су они формати и мета-подаци које је прописала нека међународна или струковна организација која обично и надгледа њихову примену.“¹⁵²

Управо из наведених разлога, постоји више схема за метаподатке које се примењују у различитим дигиталним библиотекама, а првенствено зависе од садржаја

¹⁴⁸Tim Berners-Lee. *Metadata Architecture* <http://www.w3.org/DesignIssues/Metadata.html> (приступљено 08.07.2014)

¹⁴⁹Александра Вранеш и Љиљана Марковић. *Од рукописа до библиотеке*. Метаподатак. (Београд: Филолошки факултет, 2008), 205.

¹⁵⁰Jane Greenberg. „Metadata and World Wide Web.“ *Encyclopedia of Library and Information Science*, 72, suppl. 35(2002):246

¹⁵¹Цветана Крстев. „Дигиталне библиотеке – разграничење појмова.“ *Инфотека* 3, 1-2(2002):9

¹⁵²Cvetana Krstev, Duško Vitas. „Informatički pogled na bibliografiju.“ U: Aleksandra Vraneš (ur.), *Srpska bibliografija danas*, str. 229-241 (Novi Sad: Matica srpska, 2008).230.

дигиталне библиотеке и њене намене. Развој схема за метаподатке зависи од два битна чиниоца:

1. потреба за откривањем дигиталних извора и информација у њима;
2. лак и једноставан опис дигиталних извора.

Најважније и најчешће коришћене схеме за метаподатке биће детаљно описане у посебном одељку.

Мишљења смо да термин метаподатак у домаћој библиотечко-информатичкој заједници још увек није јасно дефинисан и да има доста недоследности у погледу његовог коришћења. Библиотекари који не раде на пословима каталогизације, појам метаподатак искључиво везују за дигиталне објекте и њихов опис, док је библиотекарима каталогизаторима јасно да је то још једна врста каталогизације (особен каталошки опис) и да се може применити и на аналогне и на дигиталне објекте. Како се каталогизација донекле сматра архаичним библиотечким послом, рад са метаподацима представља се као нови изазов и значајан помак у односу на традиционалну библиотеку иако добро обучени каталогизатори аналогних објеката веома брзо овладају техником доделе метаподатака дигиталним објектима.

Као и Присила Каплан¹⁵³ сматрамо да се појам метаподатак може применити на све типове носилаца информација јер метаподаци служе да опишу објекат и омогуће његово проналажење постављањем упита за претрагу. Како је тема ове докторске дисертације проналажење информација у дигиталним библиотекама, тако ћемо се и у аспекту дескриптора метаподатака у даљем тексту бавити првенствено анализом дигиталних објеката.

4.2.2 Врсте метаподатака

Да би се представиле различите особине дигиталних објеката, користе се и различити метаподаци:

- описни или библиографски¹⁵⁴;
- административни;
- структурални.¹⁵⁵

¹⁵³Priscilla Caplan. *Metadata Fundamentals for All Librarians* (Chicago: American Library Association, 2003), 2

¹⁵⁴ У литератури се називају и интелектуалним метаподацима – Roy Tennant. „Column Digital Libraries: 21st-Century Cataloguing.” *Library Journal* (April 15, 1998), <http://roytennant.com/column/?fetch=data/96.xml> (приступљено 09.07.2014)

¹⁵⁵Подела је према Гејл Хоџ. Исту поделу користи Рој Тенант у наведеном раду.

Описни метаподаци

Када говоримо о описним (дескриптивним) или библиографским метаподацима, треба имати у виду, да се ова врста метаподатака користила још и у лисним каталозима, па после тога у електронским каталозима, само је за ту сврху коришћен другачији термин – библиографски опис. Описни метаподаци у каталозима омогућавају проналажење информација о описаном извору, које се односе на његове суштинске особине.

Описни метаподаци, из наведених разлога, наслеђе су традиционалне библиотеке јер њихов унос и врста стоје у корелацији са коришћењем стандардних формата за опис публикација још из доба лисних каталога као што је структура података у Међународном стандардном библиографском опису (ISBD), али и формата за опис публикација као што је MARC и UNIMARC, а који се јављају на преласку од традиционалне ка електронској библиотеци. Описни метаподаци су, по свом садржају, формату и употребној вредности, најсличнији библиографским подацима у системима за каталогизацију.

У дигиталним библиотекама описни метаподаци идентификују дигитални објекат и омогућавају његово проналажење. Како су описни метаподаци везани за суштинске, односно објективне особине дигиталног објекта¹⁵⁶, они укључују следеће елементе – наслов, аутор, креатор дигиталног објекта, издавач, место, година, језик, јединствени идентификациони број, опис, кључне речи, предметне одреднице, апстракт и сл. Описни метаподаци дају могућност уноса података на различитим језицима и тиме испуњавају функцију вишејезичности дигиталне библиотеке. У циљу проналажења информација, описни метаподаци дигиталних библиотека на језицима с мањим бројем говорника какав је и српски, уносе се упоредо и на енглеском језику чиме се поспешује и проналажење дигиталних објеката од стране корисника и већа видљивост српског (и других језика с мањим бројем говорника) на интернету.

Административни метаподаци

Дигитални објекат треба да буде описан и административним метаподацима јер они указују на начин управљања објектом. Саставни део административних метаподатака су и:

- метаподаци о правима коришћења (енг. rights metadata);

¹⁵⁶У литератури се среће и термин формалне особине

- технички метаподаци (енг. technical metadata) који укључују и метаподатке о очувању дигиталног објекта (енг. preservation metadata);
- метаподаци о коришћењу (енг. use metadata).

Метаподаци о правима коришћења везани су за управљање правима приступа дигиталном објекту и у складу су са ауторским правима и заштитом интелектуалне својине. Закон о ауторском и сродним правима наше земље прописује да ауторска права престају да важе 70 година након смрти аутора.¹⁵⁷ Овај податак треба имати у виду када се креира дигитална библиотека и постављају дигитализовани садржаји. Приступ комплетном тексту докумената може се омогућити само уз строго поштовање ауторских права. Аутори који су још увек активни, могу лично дозволити коришћење својих радова на различитим нивоима – делимичан приказ, само читање, али не и преузимање, приступ до нивоа сажетка и сл. Свака од поменутих ситуација мора да буде јасно дефинисана административним метаподацима. Паралела са традиционалном библиотеком може се констатовати у забрани коришћења одређених публикација ван зграде библиотеке или коришћење нпр. докторских дисертација уз потписану изјаву.

Технички метаподаци доносе податке о томе када је извор настао (датум креирања), о техничким детаљима о извору, величини и врсти датотеке, приступу извору, податке о свим изменама, о опсегу, формату приказа. Све ове информације, као и редовно освежавање промена насталих било у метаподацима, било у самом дигиталном објекту омогућавају очување дигиталног објекта и обезбеђују његово дугорочно коришћење. О значају очувања дигиталних садржаја довољно говори податак о постојању Заједнице за очување дигиталних садржаја¹⁵⁸ која доноси Препоруке за имплементацију метаподатака о очувању дигиталних објеката.¹⁵⁹

Метаподаци о коришћењу односе се на ниво и врсту коришћења дигиталног објекта – активно праћење броја корисника који посећују и користе одређени садржај, као и праћење употребе садржаја дигиталног објекта у новом контексту или у новој верзији путем преузимања метаподатака и дигиталног објекта за другу дигиталну библиотеку.

¹⁵⁷ Република Србија. „Закон о ауторском и сродним правима.” *Службени гласник Републике Србије*, бр. 104(2009), бр. 99(2011), бр. 119(2012); чл. 103, 104 и 105.

¹⁵⁸ Digital Preservation Coalition – DPC, доступно на: <http://www.dpconline.org/>

¹⁵⁹ PREMIS – Preservation Metadata: Implementation Strategies, доступно на: <http://dx.doi.org/10.7207/twr13-03>

Структурални метаподаци

Структурални метаподаци указују на структуру сложенијих дигиталних објеката. Као што је у поступку каталогизације аналогног, штампаног, извора могуће описати било цео извор, било чланак у њему или фотографију у чланку, тако и метаподаци могу да опишу дигиталне објекте на било ком нивоу агрегације.

Увек треба поћи од размишљања о томе која информација би за корисника била од највећег значаја, па се сходно томе и описују саставни делови дигиталних објеката – поглавље књиге, чланак из новина или часописа, фотографија или илустрација у оквиру неке публикације, компоненте мултимедијалних докумената (посебни видео или аудио записи, анимације или пропратни текстови). За различите дигиталне библиотеке, у зависности од врсте публикација које садрже, дефинишу се и различити нивои структуралних метаподатака.

Структурални метаподаци описују и типове, верзије, везе између дигиталних објеката и друге особине дигиталних материјала, а такође и омогућавају везе ка метаподацима надређених дигиталних објеката чији су саставни део.¹⁶⁰ Ово би значило да структурални метаподаци омогућавају да се повежу нпр. оригинална фотографија и све њене верзије при чему структурални метаподаци треба да укључе информације о верзијама, последњим променама и сл. као и везу ка извору у коме је фотографија пронађена (мета-објекту – према Армсу и др.).

Описи структуре дигиталних објеката треба да буду у складу са Функционалним захтевима за библиографске записе (FRBR¹⁶¹) које прописује Међународна федерација библиотечких удружења и институција (IFLA¹⁶²) што подразумева да се једним слогом метаподатака може описати ма који део извора – цело дело, његова верзија, поједино поглавље или чланак.

Како поступак каталогизације у традиционалној библиотеци доприноси квалитетнијем проналажењу информација, тако и описни, административни и структурални метаподаци доприносе проналажењу дигиталних објеката и информација о њима према различитим критеријумима, потом идентификују извор,

¹⁶⁰William Y. Arms, Christophe Blanchi, Edward A. Overly. „An Architecture for Information in Digital Libraries.” *D-Lib Magazine*, 3, 2(1997), <http://www.dlib.org/dlib/february97/cnri/02arms1.html> (приступљено 14.07.2014)

¹⁶¹FRBR– Functional Requirements for Bibliographic Records, доступно на: <http://www.ifla.org/publications/functional-requirements-for-bibliographic-records>

¹⁶²IFLA – International Federation of Library Association and Institution, доступно на: <http://www.ifla.org/>

помажу у организовању дигиталних објеката, окупљају сличне изворе или изворе о сличним темама, врше елиминацију истих или сличних извора или дистинкцију различитих извора и указују где се све још на мрежи могу наћи информације према постављеном упиту и дају подршку за архивирање и трајно чување дигиталних објеката.

4.2.3 Формати за унос метаподатака

Као што је раније наведено, метаподаци се у виду кодираних података користе већ и у пословима каталогизације како аналогних тако и дигиталних садржаја, у библиотечким форматима као што су MARC21 и UNIMARC, углавном помоћу атрибута „елемент = вредност“ или у некој другој синтакси. Данас је од свих формата најзначајнији и најшире распрострањен Прошириви језик за обележавање (XML¹⁶³), а у широкој употреби је и Оквир за опис извора (RDF¹⁶⁴) као виши ниво синтаксе која садржи XML. Формат XML „се подједнако користи за текстуалне податке, то јест слободан текст, као и за базе података. Може се рећи да је то данас најраспрострањенији формат за размену података између различитих рачунара, оперативних система и апликативних програма.¹⁶⁵ ”Наведене синтаксе нису метаподаци већ су општи формати података који се могу применити у различитим апликацијама за означавање докумената и нису везане искључиво за библиотеке или дигиталне библиотеке, али ћемо се ми бавити њиховом применом само у поменутом домену.

Структура синтаксе XML је у својој основи хијерархијска, нуди онолико хијерархијских нивоа колико је потребно за опис документа и омогућава размену података путем рачунара. Структуру је прописао Конзорцијум светом раширене мреже (W3C¹⁶⁶) и она је флексибилна у погледу промена – корисници је могу проширивати и прилагођавати сопственим потребама. У том смислу, изузетно је погодна за све типове метаподатака којима се описују објекти у дигиталним библиотекама. XML је, као структура за размену података, у широкој употреби и у библиотечко-информационом систему српских библиотека. Треба напоменути да сам

¹⁶³XML – eXtensibleMarkupLanguage

¹⁶⁴RDF- Resource Description Framework

¹⁶⁵Cvetana Krstev, Duško Vitas. „Informatički pogled na bibliografiju.” U: Aleksandra Vraneš (ur.) *Srpska bibliografija danas*, str. 229-241 (Novi sad: Matica srpska, 2008):231.

¹⁶⁶W3C – World Wide Web Consortium. Целокупна структура XML-а доступна на: <http://www.w3pdf.com/W3cSpec/XML/2/REC-xml11-20060816.pdf>

XML не служи за слање и пренос података. Подаци се кодирају у XML структури, а преносе преко протокола (нпр. преко HTTP¹⁶⁷ протокола).

RDF је синтакса базирана на XML коју такође подржава и развија W3C. У питању је напреднија синтакса која наглашава везу између елемената података (не само хијерархијску) и та синтакса је саставни део развоја семантичког веба. Из тог разлога је RDF изузетно значајан за повезивање метаподатака из разних дигиталних библиотека са другим изворима на мрежи. Пошто је оригинално креиран као модел за метаподатке RDF је наишао на значајну употребу у библиотекама. Међу првим библиотекама које су као формат за унос података изабрале RDF су Конгресна библиотека у Вашингтону и Сечењи библиотека у Будимпешти.

У нашој библиотечкој средини, нажалост, нема наговештаја да би RDF ускоро могао постати формат за унос метаподатака. Систем који користи највећи број библиотека базиран је на софтверу који се развија у Институту информацијских знаности (IZUM) у Марибору и имплементација нових апликација, као и прелазак на нове формате и напреднија решења зависи од плана развоја софтвера у овој установи. Иако је могуће коришћење отвореног софтвера који би садржао RDF формат, кровна библиотечка установа у нашој земљи, Народна библиотека Србије, још увек није понудила имплементацију овог формата. У неколико наврата Народна библиотека Србије организовала је радионице о коришћењу RDF формата у библиотекама где су предавачи били еминентни светски стручњаци, али се даље од тога није отишло.

Приликом одабира формата (односно схеме) за унос метаподатака за дигиталну библиотеку, треба имати у виду, осим функционалности, и његову једноставност. Библиотекари су, захваљујући искуству у раду са системима за каталогизацију, у стању да врло брзо савладају начин уноса метаподатака и особине формата за унос. Међутим, изузетно је важно да и корисници буду у стању да сами уносе метаподатке и постављају дигиталне објекте, јер поједини пројекти израде дигиталних библиотека укључују волонтерско ангажовање корисника.¹⁶⁸

¹⁶⁷HyperText Transfer Protocol

¹⁶⁸ Такви су нпр. пројекти Аустралијске националне библиотеке – Australian Newspapers Digitization Program (коректура текста) и Picture Australia (додавање и описивање слика), затим Пројекат Гутенберг – Distributed Proofreaders (креирање електронских књига), пројекат BBC-ја – World War 2 Peoples War (креирање садржаја). Занимљиви су и пројекти вишејезичних дигиталних библиотека који укључују ангажовање корисника – пројекат Конгресне библиотеке – Meeting of Frontiers (материјали на енглеском и руском језику), пројекат Latin American Open Archives Portal (материјали на шпанском и енглеском језику), као и пројекат International Children's Digital Library (материјали на енглеском, шпанском и монголском).

Ипак, и стручњаци установе која је креатор дигиталне библиотеке треба да буду изузетно опрезни у дефинисању и одабиру обавезних поља метаподатака (енг. mandatory fields). Мали број обавезних метаподатака о дигиталном објекту отежаће његово проналажење и јасно идентификовање, а превелики и захтеван смањити мотивацију корисника да раде на стварању и проширењу дигиталне библиотеке.

Формати за унос метаподатака углавном су креирани тако да су оријентисане ка корисницима, да им је структура лака за разумевање, да нису превише захтевни и да су у многоме једноставније од формата за библиографски опис публикација у традиционалним библиотекама. Ове особине показаћемо на примерима формата за унос метаподатака које су данас у најширој употреби. Неки од формата су због све веће употребе код ширег круга корисника додатно проширени и дефинисани, па су прерасли у стандарде.

Даблинско језгро

Даблинско језгро (енг. Dublin Core) је стандард за метаподатке или схема за унос метаподатака, а састоји се од сета елемената за опис широког спектра извора на мрежи. Назив потиче од места Даблин у Охају где се 1995. године одржавала иницијативна радионица под називом OCLC/NCSA Metadata Workshop. Развојем Даблинског језгра руководи Даблинска иницијатива.¹⁶⁹ Из ових основних података очигледно је да Даблинско језгро долази из библиотечке традиције. Било је доста напора да се Даблинско језгро одвоји од употребе само у библиотекама и да почне да се користи за опис докумената на интернету.¹⁷⁰

Први циљ Даблинског језгра било је дефинисање скупа елемената које би аутори веб извора користили за опис својих дела. Пошто је број дигиталних извора растао великом брзином и пошто библиотечки стручњаци нису били у могућности да каталогизују све ове изворе, прибегло се дефинисању неколико елемената и једноставних правила који би омогућили да особе које нису професионални каталогизатори постављају садржаје на мрежу и описују их. Овај принцип допринео је развоју самоархивирања као изузетно битног елемента у изради дигиталних библиотека. Остали циљеви ове иницијативе су:

- Универзално разумљива семантика и стандардизована синтакса (XML, RDF);

¹⁶⁹DCMI – Dublin Core Metadata Initiative <http://dublincore.org> (приступљено 17.8.2015)

¹⁷⁰Karen Coyle. „Understanding Metadata and Its Purpose.” *Journal of Academic Librarianship*, 31, 2(2005):162

- Међународно учешће и локализација – оригинално је писан на енглеском језику, али данас постоји и на другим језицима;
- Прилагодљивост – прилагођавање новим дигиталним изворима информација које је потребно описати, проширивање зависно од локалних потреба, као и прилагођавање у свим новим ситуацијама.¹⁷¹

Даблинско језгро се састоји од 15 основних елемената: наслов (dc:title), предмет (dc:subject), опис (dc:description), врста (dc:type), извор (dc:source), однос (dc:relation), покривеност (dc:coverage), аутор или креатор (dc:creator), издавач (dc:publisher), сарадник (dc:contributor), права (dc:rights), датум (dc:date), формат (dc:format), идентификатор (dc:identifier), језик (dc:language).

Навешћемо пример употребе Даблинског језгра за дигитални објекат на српском језику:

```
<dc:collection xmlns:dc="http://purl.org/dc/elements/1.1/"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xmlns:dcterms="http://purl.org/dc/terms/">
  <dc:dc>
    <dc:title>Razvoj geološkog terminološkog rečnika : GeolISSTerm / Ranka Stanković...[i dr.].</dc:title>
    <dcterms:alternative>Parallel title: The development of the GeolISSTerm terminological
dictionary</dcterms:alternative>
    <dcterms:alternative>Varying form of title: GeolISSTerm</dcterms:alternative>
    <dc:creator>Stanković, Ranka, 1964- (aut, code: 07527)</dc:creator>
    <dc:creator>Trivić, Branislav. (aut, code: 06698)</dc:creator>
    <dc:creator>Kitanović, Olivera. (aut)</dc:creator>
    <dc:creator>Blagojević, Branislav. (aut)</dc:creator>
    <dc:creator>Nikolić, Velizar. (aut)</dc:creator>
    <dc:language xsi:type="dcterms:ISO639-2">srp</dc:language>
    <dc:type>monographic component part</dc:type>
    <dc:type>text</dc:type>
    <dc:type xsi:type="dcterms:DCMIType">Text</dc:type>
    <dc:date xsi:type="dcterms:W3CDTF">2011</dc:date>
    <dcterms:isPartOf xsi:type="dcterms:URI">urn:ISSN:1450-9687</dcterms:isPartOf>
    <dc:relation>In: Infoteka.– Beograd : Zajednica biblioteka univerziteta u Srbiji, 2000.– ISSN 1450-9687.–
God. 12, br. 1 (avgust 2011), str. 53-67</dc:relation>
    <dcterms:bibliographicCitation>Infoteka, Beograd : Zajednica biblioteka univerziteta u Srbiji, 2000-, God.
12, br. 1 (avgust 2011), str. 53-67</dcterms:bibliographicCitation>
    <dc:format>str. 53-67. Ilustr.</dc:format>
    <dc:description>Ostali autori sa Beogradskog univerziteta: Branislav Trivić, Olivera Kitanović, Branislav
Blagojević.</dc:description>
    <dc:description>Nasl. str. prištampanog prevoda: The development of the GeolISSTerm terminological
dictionary.</dc:description>
    <dc:description>Bibliografija: str. 66-67.</dc:description>
    <dc:subject>terminološki resursi</dc:subject>
    <dc:subject>geologija</dc:subject>
    <dc:subject>GIS</dc:subject>
    <dc:subject>geološki informacioni sistemi</dc:subject>
    <dc:subject>geološki rečnik</dc:subject>
    <dc:subject>elektronski rečnik</dc:subject>
```

¹⁷¹ Marko Milenković. „Inicijativa 'Dublin Core'.” *Pregled Nacionalnog centra za digitalizaciju*, 2(2003):70.

```

<dc:subject>terminological resources</dc:subject>
<dc:subject>geology</dc:subject>
<dc:subject>GIS</dc:subject>
<dc:subject>geological information systemgeological vocabulary</dc:subject>
<dc:subject>electronic dictionary</dc:subject>
<dc:identifier xsi:type="dcterms:URI">
http://infoteka.bg.ac.rs/PDF/Srp/2011-1/INFOTHECA_XII_1_August_53-67.pdf
</dc:identifier>
<dc:identifier>38478095</dc:identifier>
</dc:dc>
</dc:collection>

```

Даблинско језгро је замишљено као једноставан и концизан стандард за опис докумената на мрежи. Међутим, тај стандард се користи и за опис других врста материјала и за сложеније примене. Све су то разлози због којих су се издвојили корисници који су заговарали, осим минималистичког, и структуралистички приступ, семантичке разлике и проширење схеме. Ти ставови довели су до разлике између квалификованог и неквалификованог Даблинског језгра. Квалификованом Даблинском језгру додати су следећи елементи и квалификатори: публика тј. намена (audience), порекло (provenance), власници права (rights holder), метод обуке (instructional method), метод прираста (accrual method), периодичност прираста (accrual periodicity), политика прираста (accrual policy).

Данас се Даблинско језгро користи и ван библиотечке заједнице, у музејима и архивима, универзитетима, и другим институцијама културе, за прикупљање и каталогизацију података са интернета, а свакако и за описивање докумената у институционалним репозиторијума. Стандард омогућава и особама које нису библиотечки стручњаци да на брз и лак начин опишу документа која постављају и тиме их, кроз процес самоархивирања, учине јавно доступним.

MODS

Пошто је MARC формат напредан и добро развијен формат за кодирање библиографских информација, у многим библиотекама енглеског говорног подручја сматрало се да је он идеалан и као стандард за метаподатке. Међутим, MARC формат, који се заснива на стандарду ISO 2709, није развијан на XML платформи. У Конгресној библиотеци у Вашингтону развијен је формат за превођење из MARC-а у XML, али он није стекао велику подршку – MARC формат је обимнији него што су објективне потребе просечних корисника, а његово коришћење због нумеричких тагова, поља и потпоља захтева додатну обуку. Због свега тога било је потребно креирати једноставнију верзију MARC формата засновану на добрим

карактеристикама XML-а. Тако је 2002. године настала пробна верзија Схеме метаподатака за опис објекта (MODS¹⁷²), а од 2003. је у употреби од стране институција и појединаца. MODS је, заправо, креиран као компромисна верзија између захтевног и компликованог MARC формата и крајње поједностављеног Даблинског језгра. Уместо нумеричких етикета, он користи разумљиве ознаке на природном језику, а употребу појединих етикета као што је нпр. ”name – име“ (лична и корпоративна имена, као и називи конференција) детаљно разрађује.

Пример записа у MODS формату

```
<modsCollection xmlns="http://www.loc.gov/mods/v3" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:schemaLocation="http://www.loc.gov/mods/v3 http://www.loc.gov/standards/mods/v3/mods-3-3.xsd">
```

```
  <mods version="3.3">
    <titleInfo>
      <title>Razvoj geološkog terminološkog rečnika</title>
      <subTitle>GeolISSTerm</subTitle>
    </titleInfo>
    <titleInfo type="alternative">
      <title>The development of the GeolISSTerm terminological dictionary</title>
    </titleInfo>
    <titleInfo type="alternative">
      <title>GeolISSTerm</title>
    </titleInfo>
    <name type="personal">
      <namePart>Stanković, Ranka</namePart>
      <namePart type="date">1964-</namePart>
      <role>
        <roleTerm authority="marcrelator" type="code">aut</roleTerm>
      </role>
    </name>
    <name type="personal">
      <namePart>Trivić, Branislav</namePart>
      <role>
        <roleTerm authority="marcrelator" type="code">aut</roleTerm>
      </role>
    </name>
    <name type="personal">
      <namePart>Kitanović, Olivera.</namePart>
      <role>
        <roleTerm authority="marcrelator" type="code">aut</roleTerm>
      </role>
    </name>
    <name type="personal">
      <namePart>Blagojević, Branislav.</namePart>
      <role>
        <roleTerm authority="marcrelator" type="code">aut</roleTerm>
      </role>
    </name>
    <name type="personal">
```

¹⁷² MODS – Metadata Object Description Scheme, доступно на: <http://www.loc.gov/standards/mods/>

```

<namePart>Nikolić, Velizar.</namePart>
<role>
<roleTerm authority="marcrelator" type="code">aut</roleTerm>
</role>
</name>
<typeOfResource>text</typeOfResource>
<originInfo>
<place>
<placeTerm authority="marccountry" type="code">rb</placeTerm>
</place>
<issuance>monographic</issuance>
</originInfo>
<language>
<languageTerm authority="iso639-2b" type="code">srp</languageTerm>
</language>
<language>
<languageTerm authority="iso639-2b" type="code">eng</languageTerm>
</language>
<physicalDescription>
<extent>str. 53-67. Ilustr.</extent>
</physicalDescription>
<targetAudience authority="marctarget">specialized</targetAudience>
<note type="statement of responsibility">Ranka Stanković...[i dr.]</note>
<note>Ostali autori sa Beogradskog univerziteta: Branislav Trivić, Olivera Kitanović, Branislav Blagojević.</note>
<note>Nasl. str. prištampanog prevoda: The development of the GeolISSTerm terminological dictionary.</note>
<note>Srp. tekst i engl. prevod štampani su u međusobno obrnutim smerovima.</note>
<note type="bibliography">Bibliografija: str. 66-67.</note>
<subject>
<topic>terminološki resursi</topic>
<topic>geologija</topic>
<topic>GIS</topic>
<topic>geološki informacioni sistemi</topic>
<topic>geološki rečnik</topic>
<topic>elektronski rečnik</topic>
</subject>
<subject>
<topic>terminological resources</topic>
<topic>geology</topic>
<topic>GIS</topic>
<topic>geological information systemgeological vocabulary</topic>
<topic>electronic dictionary</topic>
</subject>
<relatedItem type="host">
<titleInfo>
<title>Infoteka</title>
</titleInfo>
<originInfo>
<publisher>Beograd : Zajednica biblioteka univerziteta u Srbiji, 2000-</publisher>
</originInfo>
<identifier type="issn">1450-9687</identifier>
<part>
<text>God. 12, br. 1 (avgust 2011), str. 53-67</text>
</part>
</relatedItem>
<identifier type="uri">

```

```

http://infoteka.bg.ac.rs/PDF/Srp/2011-1/INFOTHECA_XII_1_August_53-67.pdf
</identifier>
<relatedItem>
<location>
<url displayLabel="electronic resource">
http://infoteka.bg.ac.rs/PDF/Srp/2011-1/INFOTHECA_XII_1_August_53-67.pdf
</url>
</location>
</relatedItem>
<recordInfo>
<descriptionStandard>ppiak</descriptionStandard>
<recordContentSource authority="marcorg">UBSM</recordContentSource>
<recordCreationDate encoding="marc">111019</recordCreationDate>
<recordChangeDate encoding="iso8601">20130111000000.0</recordChangeDate>
<recordIdentifier source="SI-MaCOB">38478095</recordIdentifier>
<recordOrigin>Converted from MARCXML to MODS version 3.3 using MARC21slim2MODS3-3.xsl
(Revision 1.50)
</recordOrigin>
<languageOfCataloging>
<languageTerm authority="iso639-2b" type="code">srp</languageTerm>
</languageOfCataloging>
</recordInfo>
</mods>
</modsCollection>

```

METS

Стандард за креирање и пренос метаподатака (METS¹⁷³) намењен је кодирању метаподатака, и намењен је првенствено дигиталним објектима и дигиталним библиотекама, а заснован је на XML платформи. За разлику од MODS-а, његова основна намена није опис докумената у смислу каталогизације. Подржава га Канцеларија за развој мреже и MARC стандарда¹⁷⁴ при Конгресној библиотеци, као и Федерација дигиталних библиотека.¹⁷⁵

Основне карактеристике METS документа су: отвореност (није у приватном власништву нити у власништву институције), развила га је библиотечка заједница, релативно је једноставан, проширив је и модуларан. Сваки METS документ може имати седам одељака: METS заглавље – metsHdr, Описни метаподаци – dmdSec, Административни метаподаци – amdSec, Одељак докумената – fileSec, Структурална мапа – structMap, Структуралне везе – structLink и Понашање – behaviorSec.

Стандард METS има нарочиту улогу у окупљању и одржавању документа који чине један дигитални објекат јер се дигитални објекти често састоје од великог броја дигиталних докумената различите врсте. METS повезује и омогућава навигацију кроз

¹⁷³ METS – Metadata Encoding and Transmission Standard, доступно на: <http://www.loc.gov/standards/mets/>

¹⁷⁴ Network Development and MARC Standards Office, доступно на: <http://www.loc.gov/marc/ndmso.html>

¹⁷⁵ DLF – Digital Library Federation, доступно на: <http://www.diglib.org/>

више дигиталних докумената истовремено. Он, такође, укључује и техничке информације неопходне за управљање дигиталним објектима – формате, технолошке карактеристике, начине скенирања, дигиталне трансформације. METS не прописује обавезни скуп описних метаподатака који ће се унети за дигитални документ, већ креатору метаподатака оставља одлуку о томе које ће описне метаподатке унети за конкретан дигитални документ. Такође, описни метаподаци за METS се могу на једноставан начин преузети из записа у Даблинском језгру или MODS-у.

LOM

LOM – Learning Object Metadata је стандард за метаподатке којим се описују наставна грађа, учила и слични дигитални објекти чија је основна намена да кориснику (ученику, студенту, полазнику курса) пружи подршку у процесу учења. Дигитални објекти који се описују помоћу овог стандарда користе се за најразличитије видове учења – школско образовање, настава на факултетима и високим школама, али и курсеви и различите врсте обуке.

LOM стандард развила је организација IEEE¹⁷⁶ још 2002. године. Стандард се састоји из два дела. Први део стандарда (1484.12.1-2002 IEEE Standard for Learning Object Metadata) односи се на метаподатке који су подељени у девет категорија – општи подаци, животни циклус, мета-метаподаци, образовни подаци, технички подаци, ауторска права, релације, анотације и класификација. У односу на остале стандарде за метаподатке, специфичност LOM стандарда лежи у томе што су, у оквиру њега, као посебна категорија, предвиђени образовни метаподаци који омогућавају пуни, односно адекватан опис наставне грађе. Други део стандарда (1484.12.3-2005 IEEE Learning Technology Standard– XML Schema Definition Language Binding for Learning Object Metadata) односи се на XML платформу на којој је LOM кодиран.^{177,178}

Најспецифичнији елементи и вредности LOM стандарда у категорији образовних метаподатака су:

- врста интеракције – Interactivity Type (активно учење, пасивно учење,

¹⁷⁶ IEEE – Institute of Electrical and Electronics Engineers

¹⁷⁷ IEEE. *Standard for Learning Object Metadata*, <https://standards.ieee.org/findstds/standard/1484.12.1-2002.html> (приступљено 20.10.2015)

¹⁷⁸ Lorna Campbell. „Learning Object Metadata (LOM).” In: S.Ross, M.Day (eds.) *DCC Digital Curation Manual*, <http://www.dcc.ac.uk/resource/curation-manual/chapters/learning-object-metadata> (приступљено 20.10.2015)

- комбинација активног и пасивног учења);
- врста материјала за учење – Learning Resource Type (вежбања, симулације, упитници, дијаграми, слајдови, текст, испити, експерименти, предавања);
 - ниво интеракције – Interactivity Level (веома низак, низак, средњи, висок, веома висок);
 - улога крајњег корисника – Intended End User Role (наставник, аутор, ученик – студент);
 - контекст – Context (школа, високо образовање, курс, друго);
 - тежина – Difficulty (веома лако, лако, средње, тешко, веома тешко).¹⁷⁹

Данас се као стандард за образовне метаподатке све чешће користи LRMI¹⁸⁰ који је настао и конципиран на основу Даблинског језгра на чију се структуру ослања, али се у погледу образовних метаподатака надовезује на елементе и поједине карактеристике LOM-а. Стандард је настао 2011, његова кодна схема је верификована 2013, са примарним циљем да се образовним метаподацима опише наставна грађа која се налази на мрежи, а намењена је како истраживачима и предавачима, тако и ученицима, студентима и свим другим особама које су у процесу учења.

TEI

TEI – Иницијатива за кодирање текста (Text Encoding Initiative) је пројекат који је покренут 1987. са циљем да као софтверски независна метода развија и одржава податке из хуманистичких наука (нарочито у литерарним и лингвистичким истраживањима) који се депонују у електронском облику.¹⁸¹ Пројекат су покренуле асоцијације Association for Computers and the Humanities (ACH), Association for Literary and Linguistic Computing (ALLC) и Association for Computational Linguistics (ACL). Године 1994. године објављено је прво издање препорука за обележавање и размену електронских текстова Guidelines for Electronic Text Encoding and Interchange. ”Пројекат је извршио исцрпну анализу и идентификацију категорија и својстава за обележавање текстуалних података, имајући у виду разноврсност садржаја: литерарна

¹⁷⁹ Александра Трговац. „Метаподаци у систему PHAIDRA – дигиталном репозиторијуму Универзитета у Београду.” У: Весна Црногорац, Весна Ињац (ур.) *Хоризонти светског и европског библиотекарства у дигиталном добу*, Међународна научна конференција, Београд 27-28. октобар 2011, стр. 171-188. (Београд: Библиотекарско друштво Србије, 2012), , <http://bds.rs/wordpress/wp-content/uploads/2013/08/Horizonti-svetskog-i-evropskog-bibliotekarstva-u-digitalnom-dobu.pdf> (приступљено 20.10.2015)

¹⁸⁰ LRMI – Learning resources Metadata Initiative, доступно на: <http://dublincore.org/dcx/lrmi-terms/>.

¹⁸¹ TEI – Text Encoding Initiative. History, <http://www.tei-c.org/About/history.xml> (приступљено 19.10.2015)

проза, поезија, драмски текстови, речници, термилошке базе, историјски извори, и тако даље. Прве верзије препорука су као формат користиле језик SGML (Structured Markup Generalized Language), који је претходник савременијег језика за обележавање XML који користе новије верзије препорука.¹⁸²

TEI препоруке доносе исцрпну листу метаподатака које се уносе у заглавље TEI документа и објашњавају на који начин се врши записивање библиографских информација како о електронској верзији текста тако и о опису штампаног документа при чему је скуп метаподатака о електронској верзији одвоје од скупа података о штампаној верзији. Основне информације у библиографском опису електронског документа наликују на класичну библиотечку каталогизацију, али садрже и елементе који описују на који начин је текст обележен, како је транскрибован, колико пута је текст допуњен и ревидиран укључујући податке о томе ко је вршио измене и какве природе су те измене, те нуди могућност приступа свим претходним верзијама.¹⁸³

TEI има два основна смера коришћења:

- креирање дигиталних библиотека, нарочито оних које садрже доста текстуалног материјала, а посебно старе и ретке грађе која једино као дигитална може да буде доступна ширем кругу корисника (у TEI препорукама се јасно дефинише који се библиографски метаподаци уносе да би проналажење информација о таквој грађи било што боље);
- креирање мањих, тематски оријентисаних колекција у оквиру појединих истраживачких пројеката (одређени лексички корпус, књижевни жанр, аутор, историјска перспектива или географска одређеност или комбинација критеријума).¹⁸⁴

TEI препоруке су данас после три деценије коришћења у широкој употреби у оквиру значајних светских научних пројеката и институција (Cambridge University Press, Center for Electronic Texts in the Humanities, Oxford Text Archive, British National Corpus, Corpus Encoding Standard EpiDoc, FreeDict и др.).

¹⁸²Cvetana Krstev, Duško Vitas. „Informatički pogled na bibliografiju.” U: Aleksandra Vraneš (ur.) *Srpska bibliografija danas*, str. 229-241 (Novi Sad: Matica srpska, 2008):237.

¹⁸³TEI Consortium. TEI P5 – Guidelines for Electronic Text Encoding and Interchange, <http://www.tei-c.org/Vault/P5/2.9.1/doc/tei-p5-doc/en/Guidelines.pdf> (приступљено 19.10.2015)

¹⁸⁴Томаж Ерјавец. „Смернице иницијативе за кодирање текста и њихова локализација.” *Инфотека* 11, 1 (2010):5, http://infoteka.bg.ac.rs/pdf/Srp/2010-1/INFOTHECA_XI_1_April2010_3-15.pdf (приступљено 19.10.2015)

За истраживања из српског језика која су базирана на TEI препорукама изузетно је значајан пројекат MULTEXT-East.¹⁸⁵ У оквиру њега користе се језички ресурси 16 језика из средње и источне Европе, а најважнији ресурс је роман Џорџа Орвела ”1984” на оригиналном енглеском језику и преводима на српски и све остале језике који су обухваћени пројектом.¹⁸⁶ MULTEXT-East обухвата: текстове обележене структурама с TEI етикетама; етикетиране текстове који садрже лему, врсту речи, морфодинамичке категорије; текст који је поравнат са енглеским оригиналом. Још једно значајно истраживање које користи TEI препоруке за доделу библиографских података електронском тексту јесте електронско издање српских народних пословица Вука С. Караџића.¹⁸⁷

Описане схеме за метаподатке су једноставне за коришћење и омогућавају унос података, као и размену података са другим системима. Направљене су тако да кориснику пружају могућност да брзо и лако постављеном дигиталном објекту додели метаподатке неопходне за претраживање и проналажење. Од вештине корисника, зависно од тога да ли је професионалац или почетник зависи колико детаљно ће један дигитални објекат бити описан, а самим тим какав ће бити квалитет претраживања и проналажења информација о конкретном дигиталном објекту.

4.2.4 Системи за унос података о објектима

Полазећи од каталогизације у машински читљивим форматима и каталошким записима у њима ка описивању дигиталних објеката са пратећим метаподацима посебну пажњу ћемо посветити анализи записа и метаподатака у следећим системима:

- а. Узајамна библиографско-каталожка база података COBIB.SR;
- б. Институционални репозиторијуми;
- в. Еуропеана.

а. Узајамна библиографско-каталожка база података COBIB.SR

Узајамни каталог садржи библиографске податке две или више библиотека, као и локацијске податке библиотека које у том систему учествују ради идентификовања

¹⁸⁵ MULTEXT-East: Multilingual Text Tools and Corpora for Central and Eastern European Languages, доступно на: <http://nl.ijs.si/ME/>

¹⁸⁶ Tomaž Erjavec. “MULTEXT-East: morphosyntactic resources for Central and Eastern European languages.” *Language Resources and Evaluation* 46, 1(March 2012):131-142.

¹⁸⁷ Cvetana Krstev, Duško Vitas, Vesna Šatev. “TEI Encoding of Serbian Proverbs.” In: Anisava Miltenova et al. (eds.) *Computer Applications in Slavic Studies*, Proceedings of Azbuky.Net International Conference and Workshop, 24-27 October 2005, Sofia, Bulgaria, , pp. 101-114 (Sofia: Institute of Literature, Bulgarian Academy of Sciences, 2006)

података о стању фонда.¹⁸⁸ Овакав каталог треба да омогући брзу и једноставну размену записа и императив је сваког савременог библиотечког система. На овим постулатима грађен је и систем COBISS.SR.

Систем COBISS.SR (Кооперативни онлајн библиографски систем и сервис) је интегрисани библиотечки систем Србије и саставни део ширег регионалног система COBISS.Net који је састављен од библиографско-каталожких база података Словеније (442 библиотеке), Србије (171 библиотека), Босне и Херцеговине (Федерација БиХ 50 библиотека, Република Српска 15 библиотека), Македоније (42 библиотеке), Црне Горе (28 библиотека), Албаније (29 библиотека) и Бугарске (2 библиотеке).¹⁸⁹ Систем COBISS настао је, развија се и одржава у Институту информацијских знаности (IZUM) у Марибору још од 1987. године када је успостављен као систем за узајамну каталогизацију библиотека Југославије.

Узајамна библиографско-каталожка база података COBIB.SR представља резултат узајамне каталогизације српских библиотека укључених у систем COBISS.SR. Садржи библиографске записе о књижној (монографске публикације, серијске публикације, чланке) и некњижној грађи (CD, DVD и др.), као и о публикацијама (монографским, серијским, чланцима) које су објављене само на вебу или у оквиру дигиталних библиотека (дигитални објекти). База података COBIB.SR од основног је значаја за систем COBISS.SR, јер има централну улогу у процесу узајамне каталогизације, у претраживању библиографских информација, као и информација о локацији и доступности грађе у српским библиотекама.¹⁹⁰

У овој библиографско-каталожкој бази, метаподатке о публикацијама свих типова похрањује преко 160 највећих српских библиотека, укључујући Народну библиотеку Србије, Библиотеку Матице српске, Универзитетску библиотеку „Светозар Марковић“, универзитетске библиотеке у Нишу и Крагујевцу, Библиотеку Српске академије наука и уметности, факултетске библиотеке државних и приватних факултета у Београду, Новом Саду, Крагујевцу и Нишу, као и јавне (народне)

¹⁸⁸ John Feather and Paul Sturges (eds.) *International Encyclopedia of Library and Information Science*. Union Catalogues. (London: Routledge, 2003), 643.

¹⁸⁹ COBISS.Net, <http://www.cobiss.net/default-sr.asp> (приступљено 19.10.2015)

¹⁹⁰ COBIB.SR, доступно на: <http://www.vbs.rs/scripts/cobiss?ukaz=INFO&id=1946024215347051&pg=1>

библиотеке бројних градова и општина широм Србије и специјалне библиотеке различитих организација.¹⁹¹

За размену података у систему COBISS користе се формат COMARC/B за библиографске податке и формат COMARC/A за нормативне податке базирани на UNIMARC формату. За међународну размену библиографских података користи се и формат MARC 21. Записи су доступни у структурама ISO2709 (у форматима MARC21 и COMARC) или XML (формати DublinCore, MODS, MARC 21, COMARC). Извоз библиографских метаподатака могућ је у форматима DublinCore, MODS, MARC 21 XML, COMARCXML.

Од изузетне важности у погледу стандардизације и међународне размене података је чињеница да су сви ови формати у складу са међународним стандардима и омогућавају комуникацију и размену података уз минималне губитке у погледу садржаја и функционалности између различитих система, односно различитих дигиталних библиотека.¹⁹²

Овако структурирани метаподаци користе се и код класичних, штампаних публикација, али се могу користити и код описа дигиталних објеката и омогућити квалитетније проналажење информација о њима. Таква ситуација олакшава слање метаподатака о дигиталним објектима другим системима и учешће у бројним пројектима путем којих се представљају и повезују подаци о различитим дигиталним библиотекама.

У овом раду поменути формати примениће се у процесима проналажења информација у записима за различите типове грађе, као и код описаних објеката у дигиталним библиотекама, а резултати ће се упоредити са информацијама добијеним претрагом у другим анализираним системима (нарочито у онима који садрже записе за идентичне објекте).

¹⁹¹ Систем COBISS није једини библиотечки систем који се користи у српским библиотекама. Поред овог система, у употреби су и системи BISIS (Библиотечки информациони систем развијан на Универзитету у Новом Саду од 1992. године, користи се у бројним јавним (народним) и факултетским библиотекама у Војводини, али и у Библиотеци града Београда и њеним огранцима, информације доступне на: <http://bisis.uns.ac.rs/wiki/WikiStart>), затим NIBIS (Нишки библиотечко-информациони систем развијан на Универзитету у Нишу од 2002. године, користи се у јавним (народним) библиотекама, као и у већем броју музејских и специјалних библиотека, информације доступне на: <http://icit.masfak.ni.ac.rs/sitegenius/topic.php?id=157>), као и други системи који су заступљени у мањем броју библиотека (BIBLIO, PERGAM и сл.).

¹⁹² У међународну размену записа, према међународним договорима, не улазе подаци о стању фонда који се уносе у формату COMARC/H. У систему COBISS.SR тренутно је у употреби корисничко сучеље COBISS 3 за унос библиографских, нормативних и података о стању фонда.

На примеру ћемо показати изглед формата за унос библиографских података COMARC/B као и изглед формата за експорт података из истог записа у Даблинском језгру (Dublin Core) XML.

Пун	ISBD	COMARC
ID=146708748 M V3 21.02.2008 NBS::ANAS Updated: 20.08.2008 UBSM::TANJA Copied: 04.07.2008 UBSM::LALIC		
001		aс - исправљени запис bа - текстуална грађа, штампана cm - монографска публикација d0 - нема хиерархијског односа 7ba - латиница t2.02 - стручна монографија
010		a978-86-86419-43-9
100		c2008 hssc - српски lba - латиница
101	0	aeng - енглески
102		asrb - Србија bcs - Централна Србија
105		aa - илустрације b4 - стручни рад
200	0	aProcessing of Serbian eautomata, texts and electronic dictionaries fCvetana Krstev
210		aBelgrade cFaculty of Philology of the University d2008 eBelgrade gDS print
215		a228 str. cgraf. prikazi d24 cm
300	0	aTiraž 300
320	0	aNapomene i bibliografske reference uz tekst
320	0	aBibliografija: str. 219-227.
606	3	aSrpski jezik xLeksika xPrimena računara
606	3	aSrpski jezik xElektronski rečnici
675		a004.9:811.163.41374 c004.9 - Апликативно оријентисане технике засноване на рачунарима
675		c811.163.41374 - Лексикографија. Речници српског језика
675		a811.163.41322 c811.163.41 - Српски језик
686		aFillc2c2 2UBBG - Стручни каталог Универзитетске библиотеке "СВЕТОЗАР МАРКОВИЋ", Београд
700	1	aKrstev bCvetana 4070 - аутор f1952- 700006 8SR13-01.22 - Универзитет у Београду, Филолошки факултет

Слика 1 – Пример записа за монографску публикацију у формату COMARC/B узајамног каталога COBIJ.SR

This XML file does not appear to have any style information associated with it. The document tree is shown below.

```

▼<dc:collection xmlns:dc="http://purl.org/dc/elements/1.1/" xmlns:xsi="http://www.w3.org/2001/XMLSchema-
instance" xmlns:dcterms="http://purl.org/dc/terms/"
▼<dc:dc>
▼<dc:title>
Processing of Serbian : automata, texts and electronic dictionaries / Cvetana Krstev.
</dc:title>
<dc:creator>Krstev, Cvetana. 1952-(aut, code: 00006)</dc:creator>
<dc:type>monograph</dc:type>
<dc:type>text</dc:type>
<dc:type xsi:type="dcterms:DCMIType">Text</dc:type>
▼<dc:publisher>
Belgrade : Faculty of Philology of the University,
</dc:publisher>
<dcterms:issue>2008</dcterms:issue>
<dc:date xsi:type="dcterms:W3CDTF">2008</dc:date>
<dc:format>228 str. : graf. prikazi ; 24 cm.</dc:format>
<dc:description>Tiraž 300.</dc:description>
<dc:description>Napomene i bibliografske reference uz tekst.</dc:description>
<dc:description>Bibliografija: str. 219-227.</dc:description>
<dc:subject>Srpski jezik</dc:subject>
<dc:subject>Srpski jezik</dc:subject>
<dc:identifier xsi:type="dcterms:URI">urn:ISBN:978-86-86419-43-9</dc:identifier>
<dc:identifier>146708748</dc:identifier>
</dc:dc>
</dc:collection>

```

Слика 2 – Пример записа за монографску публикацију у формату за експорт података Dublin Core XML

б. Институционални репозиторијуми

Према Клифорду Линчу институционални репозиторијум је „скуп услуга које институција, а најчешће универзитет нуди својим члановима за управљање и дистрибуцију дигиталних материјала, укључујући и њихову дугорочну заштиту.”¹⁹³ Марк Вејр дефинише институционални репозиторијум као „базу података засновану на веб технологији, а која садржи материјале који су дефинисани профилем конкретне установе (а не садржај самог материјала).“¹⁹⁴

Појам институционалног репозиторијума шире дефинише и објашњава Кроу¹⁹⁵ који истиче да институционални репозиторијум служи креирању не само креирању дигиталних збирки већ и очувању и интелектуалном утицају једне или више универзитетских заједница.

На основу претходно наведених дефиниција може се закључити да је дигитални институционални репозиторијум локација на мрежи на коју се постављају колекције дигиталних докумената, врши њихова заштита (у смислу ауторских права) и дисеминација и давање на коришћење под одређеним, јасно прописаним условима. Универзитетски репозиторијуми су институционално и тематски организоване дигиталне збирке докумената и података о научном, истраживачком и едукативном раду на универзитету.

Дигитални репозиторијум универзитета у својим дигиталним колекцијама сакупља радове наставника и сарадника, као и студената конкретног универзитета. У саставни део тих дигиталних колекција спадају: уџбеници релевантни за студије на универзитету, научни часописи и монографије, чланци из научних зборника и периодике, препринтови и постпринтови научних чланака, дигиталне верзије теза и дисертација брањених на универзитету, разна административна документа, разне врсте учила значајних за образовање на универзитету.

У складу са Кроуовом дефиницијом, основне улоге дигиталног репозиторијума универзитета су:

¹⁹³ Clifford A. Lynch. „Institutional Repositories: Essential Infrastructure for Scholarship in the Digital Age.” *ARL* 226(2003):2, , <http://www.arl.org/bm~doc/br226ir.pdf> (приступљено 19.08.2014)

¹⁹⁴ Mark Ware. *Pathfinder Research on Web-based Repositories*. (London: Publisher and Library/Learning Solution, 2004), 3, <http://mrkwr.files.wordpress.com/2006/11/pals-report-on-institutional-repositories.pdf> (приступљено 19.08.2014)

¹⁹⁵ Raym Crow. *The Case for Institutional Repositories: A SPARC Position Paper*. (Washington, DC: The Scholarly Publishing and Academic Resources Coalition, 2002), 4, http://www.arl.org/sparc/bm~doc/ir_final_release_102.pdf (приступљено 19.08.2014)

1. Отворени приступ резултатима истраживања на универзитету и њихово самоархивирање;
2. Глобална видљивост научног рада и истраживања на универзитету;
3. Обједињавање дигиталних садржаја значајних за универзитет на једној веб локацији;
4. Похрањивања и чување и других дигиталних објеката укључујући необјављене радове (тзв. „сиву“ литературу, тезе и техничке извештаје).

У овој докторској дисертацији анализираћемо структуру, садржај, метаподатке и доступност дигиталних објеката у два институционална репозиторијума: PHAIDRA – Дигитални репозиторијум Универзитета у Београду и Дигитална библиотека Филолошког факултета.

PHAIDRA

PHAIDRA¹⁹⁶ је систем за дугорочно похрањивање, индексирање и коришћење дигиталних објеката на Универзитету у Београду. То је званични репозиторијум Универзитета у Београду кога одржава Универзитетска библиотека „Светозар Марковић.” Према основим концептима, предвиђено је да овај репозиторијум омогући да се резултати научно-истраживачког, теоријског, практичног и уметничког рада на Универзитету у Београду систематски архивирају, трајно чувају и учине доступним јавности.¹⁹⁷ Репозиторијум је успостављен у оквиру Темпус пројекта „Нови библиотечки сервиси на универзитетима Западног Балкана“ (2011-2013)¹⁹⁸ и, осим на Универзитету у Београду, имплементиран је на универзитетима који су били партнери у Пројекту – у Нишу, Крагујевцу, Подгорици, Сарајеву и Тузли.

Репозиторијум PHAIDRA чува и трајно архивира научне, стручне, прегледне, докторске, мастер, дипломске и друге публиковане и непубликоване радове уз потпуно поштовање ауторских права са тенденцијом ка отвореном приступу. У њега се могу похранити дигитални објекти различитих врста (текстови, слике, аудио и видео записи, мултимедијални документи) који прате наставни, научни, административни и организациони рад на Универзитету. Сви објекти се описују

¹⁹⁶Permanent Hosting, Archiving and Indexing of Digital Resources and Assets. Репозиторијум је иницијално развијен и имплементиран на Универзитету у Бечу.

¹⁹⁷PHAIDRA – Дигитални репозиторијум Универзитета у Београду, доступно на:
<https://phaidrabg.bg.ac.rs/>

¹⁹⁸ Темпус пројекат „Нови библиотечки сервиси на универзитетима Западног Балкана“, доступно на:
<http://www.westbulnet.com/AmplioCMS2/public/EnterPublic.cms2#menuId=1311>

стандаризованим метаподацима и сваки од објеката након похрањивања добија трајни линк што омогућава његово лакше лоцирање и цитирање.

Дигитални објекти се могу постављати у виду аудио и видео записа, књига, колекција, контејнера који садржи различите колекције, текстуалних LaTeX и pdf докумената, слика и некатегоризованих докумената. Колекције и контејнери су предвиђени за посебне тематске збирке, за бројеве једног наслова часописа, радове у оквиру једног зборника и сл.

Да би се документа похрањивала у репозиторијум PHAIDRA неопходан услов је да буду дигитализована и преведена у следеће препоручене стандардне формате:

- текстуални подаци конвертују се у PDF у оквиру ког је помоћу програма за оптичко препознавање карактера (OCR) могуће претраживање пуног текста;
- сликовна грађа преводи се у формате TIFF, JPEG;
- видео записи дигитализују се у форматима AVI, MPEG2;
- аудио записи имају препоручени формат WAV.

Метаподаци

Систем PHAIDRA на свим универзитетима где је до сада постављен нуди више могућности за унос метаподатака. Даблинско језгро је генерисано у самом систему и то је основна понуђена могућност. Такође, XML ток података сачуван је за сваки објект који се налази у репозиторијуму. PHAIDRA нуди модификовану схему за метаподатке LOM¹⁹⁹, али подржава и стандард за метаподатке METS. Универзитетима и факултетима је омогућено да направе схему за метаподатке у складу са својим потребама, а она лако може бити имплементирана у репозиторијум PHAIDRA.²⁰⁰ На Универзитету у Београду, а на предлог Универзитетске библиотеке „Светозар Марковић“, одлучено је да се као иницијална схема за метаподатке користи Даблинско језгро.

¹⁹⁹LOM – Learning Object Metadata, доступно на: <http://standards.ieee.org/findstds/standard/1484.12.1-2002.html>

²⁰⁰University of Vienna. *Phaidra: Permanent Hosting, Archiving and Indexing of Digital Resources and Assets*, http://www.westbulnet.com/AmplioCMS2/public/EnterPublic.cms2?langId=74&request_locale=sr_CS#menuId=1698 (приступљено 20.08.2014)

Identifikator*:	o:5568	
Naslov*:	Biblioteke verskih zajednica i institucija	na jeziku: srpski
Podnaslov:	doktorska disertacija	na jeziku: srpski
Uporedni naslov:	Libraries of religious communities and insi	na jeziku: engleski
Externalview:		
Jezik*:	srpski	
Opis ili dodatni podaci*:	Bibliotekarstvo i informatika - Specijalne biblioteke Datum odbrane: 24.04.2013.	na jeziku: srpski
	Dodajte link(ove)	
Opis ili dodatni podaci*:	Library an information science - Special libraries	na jeziku: engleski
	Dodajte link(ove)	
Ključne reči:	BIBLIOTEKE VERSKIH ZAJEDNICA I IN	na jeziku: srpski
Ključne reči:	RELIGIOUS LIBRARIES, RELIGIOUS LI	na jeziku: engleski
Opseg:		na jeziku: srpski
Institucionalni repozitorij:	da	
Identifikatori		
Izvor:	(bez izbora)	
Identifikator:	43505679	

Слика 3 – Општи библиографски метаподаци у формату за унос репозиторијума PHAIDRA

Предвиђена је могућност уноса следећих описних, односно дескриптивних (библиографских) метаподатака: наслов, креатор, афилијација (податак о институцији у оквиру које аутор делује), апстракт, дескриптори, датум, идентификатор, држава издавања, јединствени идентификациони бројеви (ISSN, ISBN, ISMN), језик, издавач, место издавања, година издавања, опсег документа, годиште, број, наслов надређене публикације.²⁰¹ Део ових метаподатака приказан је на сликама 3 и 4.

²⁰¹ Темпус пројекат „Нови библиотечки сервиси на универзитетима Западног Балкана“, доступно на: http://www.westbulnet.com/AmplioCMS2/public/EnterPublic.cms2?langId=74&request_locale=sr_CS#menuId=1347

The image shows two screenshots of a web form for entering data into the PHAIDRA repository. The top screenshot shows the 'Doprinosi' (Contributions) section with the role 'autor digitalnog objekta' (author of digital object). The 'Entitet / lični podaci*' (Entity / personal data) section contains the following fields: 'Ime' (Name) with 'Nevena, 197', 'Prezime' (Surname) with 'Tomić', and two empty 'Akademsko/stručno zvanje' (Academic/professional title) fields. The 'Datum' (Date) field is set to '(DD.MM.YYYY):' with a dropdown menu showing 'nema podataka' (no data) and the year '2012'. The bottom screenshot shows the same 'Doprinosi' section but with the role 'drugo' (other) and the 'Različita uloga' (Different role) field set to 'mentor'. The 'Entitet / lični podaci*' section contains: 'Ime' with 'Aleksandra, 1', 'Prezime' with 'Vraneš', and two empty 'Akademsko/stručno zvanje' fields.

Слика 4 – Подаци о аутору и ментору докторске дисертације у формату за унос података у репозиторијуму PHAIDRA

Административни метаподаци (видети слику 5) распоређени су у оквиру репозиторијума PHAIDRA у неколико категорија:

- Тип документа: докторски, магистарски, спрецијалистички, дипломски рад, издавачка продукција универзитета, монографије, чланци и прилози у серијским публикацијама, објављено саопштење на научном скупу, препубликовани документ, пројектни документ, институционални документ, слике, видео записи, веб презентације, тонски записи, наставна грађа;
- Верзија документа;
- Институционално власништво: универзитет, факултет, катедра, смер, одељење;
- Персонални власник документа;
- Дозвољени приступ документу: власник, факултет, катедра, смер, одељење, запослени на факултету, без ограничења, као и ограничавање могућности да се користи пун текст документа.

Предвиђени метаподаци за заштиту и коришћење докумената:

- Врста формата документа;
- Услови коришћења документа;
- Датум првог похрањивања у репозиторијум;
- Датум последње промене података;

- Величина датотеке.

Kao autor ili vlasnik prava, imate mogućnost da date ovlašćenje za korišćenje autorskog dela, za umnožavanje, distribuciju, emitovanje, dopunu i izmenu zaštićenog dela korišćenjem Creative Commons licenci. Zadržavate svoja autorska prava. Možete da precizirate druge uslove kako biste zaštitili svoja dela izborom licence koja odgovara vašim željama.

Troškovi*:

Копирајт и друга ограничења*:

Izabrana licenca*:

Licence - kraća verzija

Dozvoljavate umnožavanje, distribuciju i javno saopštavanje dela, i prerade, ako se navede ime autora na način odredjen od strane autora ili davaoca licence, čak i u komercijalne svrhe. Ovo je najslobodnija od svih licenci.

Osnovni opis Licence:

http://creativecommons.org/licenses/by/3.0/rs/deed_sr_LATN

Слика 5 – Административни метаподаци о врсти лиценце за коришћење дигиталног објекта у формату за унос података у репозиторијуму PHAIDRA

Неопходно је истаћи да је у аустријском репозиторијуму PHAIDRA развијен посебан део који се бави применом класификационих система у оквиру уноса метаподатака (ÖFOS²⁰², EuroVoc²⁰³, и други сходно потребама универзитета и факултета).

Klasifikacija

Klasifikacije (klase, podklase)

Podrška za pretraživanje

Izvor:

Putanja*:

Opis ili dodatni podaci: **na jeziku:**

Dodajte link(ove)

Ključne reči: **na jeziku:**

Ključne reči: **na jeziku:**

Ključne reči: **na jeziku:**

Ključne reči: **na jeziku:**

Слика 6 – Библиографски метаподаци о класификацији у формату за унос података у репозиторијуму PHAIDRA

²⁰² Österreichische Systematik der Wissenschaftszweige, доступно на: www.statistik.at/kdb/pages/info_10515.html

²⁰³ Multilingual Thesaurus of the European Union, доступно на: <http://eurovoc.europa.eu/drupal/>

За репозиторијум универзитета у Србији осим тезауруса EuroVoc који постоји и на српском језику, у Универзитетској библиотеци „Светозар Марковић“ израђена је и класификациона схема ОСНО (Основна схема научних области) која се ослања на УДК систем, а исписује се текстуално, не нумерички (видети слику 6).

Од свих наведених метаподатака, обавезан је унос само осам елемената приликом постављања дигиталног објекта: наслов, језик, опис, статус, улога (ауторство), цена (трошкови), ауторска права и друга ограничења, лиценца (власништво ауторских права). Датум и идентификатор се аутоматски генеришу.

Због природе дигиталних објеката, осим стандардних метаподатака (описних, административних и структуралних), дигитални документи у овом Репозиторијуму су описани и посебним метаподацима везаним за наставу, наставну грађу и учила. У ту сврху коришћен је стандард LOM (Learning Object Metadata).

Овако дефинисани параметри за унос у многа олакшавају процес постављања дигиталних објеката у репозиторијум РНАИДРА и омогућавају и особама које нису стручњаци за каталогизацију да то самостално учине, што је и смисао оваквих комуникационих система. Ипак, са друге стране, библиотекари који раде на уносу објеката се труде да метаподаци буду што детаљнији да би одзив приликом проналажења информација био квалитетнији.

Због боље видљивости дигиталних објеката за докторске дисертације, метаподаци из овог репозиторијума су, преко портала doiSerbia PhD²⁰⁴ на коме свака докторска дисертација добија нови метаподатак – DOI број, повезани са европским порталом докторских дисертација – DART Europe.²⁰⁵

²⁰⁴ doiSerbia PhD, доступно на: <http://www.doiserbia.nb.rs/phd/Default.aspx>

²⁰⁵ DART Europe, доступно на: <http://www.dart-europe.eu/basic-search.php>

DART-Europe E-theses Portal

HOME ABOUT DART-EUROPE DOCUMENTS & DOWNLOADS RELATED ORGANISATIONS PARTNERS & BOARD

Full record Search Browse Search History Marked List Results Full Record Feedback Help

Search again Share this record Save this record [« Back to results](#)

Search term(s) Search

Showing record 2 of 21 [« Previous record](#) [Next record »](#)

Title	Libraries of religious communities and institutions
Title	Biblioteke verskih zajednica i institucija
Publisher	University of Belgrade, Faculty of Philology
Type	PhD thesis
Format	application/pdf
Author	Tomić Nevena
Contributor(s)	Tanasković Darko; Vraneš Aleksandra; Šijaković Bogoljub
Subject(s)	Biblioteke verskih zajednica i institucija, Libraries of religious communities, Religijske biblioteke, Religijsko bibliotekarstvo, Religious librarianship, Religious libraries, Verske biblioteke
Abstract	This dissertation deals with the religious libraries - libraries of religious communities and institutions in the world. It tends to give a picture of theory and practice in this special field of librarianship. Although these libraries have long history this scientific field is very young and still in development. Brief overview of theoretical works and used terminology in several world languages is given. Historical development of libraries in traditional religions of East and West shows their importance for the spread of religions and culture. The social role of the religious library in the frame of the religious and local community is also considered. Intellectual freedom, censorship and ethical issues are particularly related to libraries which belong to religious communities, because of their mission to promote ethical attitudes and religious beliefs of the community. Special challenges of cataloging and classification in the field of theology and religion are discussed. Considering collection development the focus is given to the good examples of library consortia that exist in the field. Professional associations and their activities are presented. Digital environment raises questions of the adequate presence of the religious libraries on the Internet, OPAC, resources in the field of religion and theology, digitalization projects in the field. Four different cases (Family History Library in Salt Lake City, Library of Tibetan Works and Archives in India, Christian Science Reading Rooms and Chilandar Scientific Library in Ohio) are presented as examples of the variety of religious libraires.

Abstract	Ova disertacija bavi se bibliotekama verskih zajednica i institucija u svetu, sa ciljem da prikaže razvoj i domete teorije i prakse u ovoj posebnoj oblasti bibliotekarstva. Iako ove biblioteke imaju dugu istoriju, sama naučna oblast je mlada i još uvek se razvija. Na početku rada dat je prikaz teorijskih radova i korišćene terminologije na nekoliko svetskih jezika. Istorijski razvoj biblioteka verskih zajednica i institucija zapadnih i istočnih tradicionalnih religija ukazuje na važnost njihove uloge u širenju religija i kultura. Razmatrana je društvena uloga biblioteka verskih zajednica i institucija u okvirima verske i društvene zajednice. Pitanja intelektualne slobode, cenzure, etike, posebno su značajna za ove biblioteke zbog njihove misije da promovišu etičke stavove i religijska verovanja svoje zajednice. Ovaj rad bavi se i pitanjima prakse u bibliotekama verskih zajednica: katalogizacijom, klasifikacijom, selekcijom, nabavkom i razvojem fondova, radom sa korisnicima. Katalogizacija i klasifikacija u oblastima religije i teologije su specifične oblasti koja zahtevaju visokostručna znanja. Opisani su i primeri dobre prakse konzorcijuma u oblasti objedinjene nabavke i razvoja fondova. Predstavljena je aktivnost profesionalnih udruženja biblioteka verskih zajednica i institucija koja postoje u svetu. Digitalno okruženje pokreće pitanja adekvatnog predstavljanja na Internetu, Web kataloga, digitalnih izvora u oblastima religije i teologije, projekata digitalizacije koji postoje u ovoj oblasti. Kao primeri specifične prakse i raznovrsnosti koja postoji među bibliotekama verskih zajednica i institucija opisani su: Biblioteka porodične istorije koja pripada mormonskoj zajednici, nacionalna Biblioteka tibetskih radova i arhiva u Indiji, čitaonice verske zajednice Hrišćanska nauka i Hilendarska naučna biblioteka koja se nalazi na Državnom univerzitetu u Ohaju.
Date	2013-04-24
Identifier	http://dx.doi.org/10.2298/BG20130424TOMIC
Identifier	http://eteze.bg.ac.rs/application/showtheses?thesesId=274
Identifier	https://fedorabg.bg.ac.rs/fedora/get/o:5568/bdef:Content/get
Identifier	http://vbs.rs/scripts/cobiss?command=SEARCH&base=99999&select=ID=43505679
Language	sr

Слика 7 – Пример записа за исту докторску дисертацију са припадајућим метаподацима на порталу европских дисертација DART

На слици 7 представљен је пример записа за домаћу докторску дисертацију на порталу DART Европе, а на слици 8 запис за исту докторску дисертацију у формату за експорт података.

This XML file does not appear to have any style information associated with it. The document tree is shown below.

```

▼<oai_dc:dc xmlns:dc="http://purl.org/dc/elements/1.1/" xmlns:oai_dc="http://www.openarchives.org/OAI/2.0/oai_dc/">
  <dc:creator>Tomić, Nevena, 1972- (aut)</dc:creator>
  <dc:language>sr</dc:language>
  ▼<dc:subject>
    BIBLIOTEKE VERSKIH ZAJEDNICA I INSTITUCIJA, RELIGIJSKE BIBLIOTEKE, VERSKE BIBLIOTEKE, RELIGIJSKO BIBLIOTEKARSTVO
  </dc:subject>
  ▼<dc:subject>
    RELIGIOUS LIBRARIES, RELIGIOUS LIBRARIANSHIP, LIBRARIES OF RELIGIOUS COMMUNITIES
  </dc:subject>
  ▼<dc:subject>
    OSNO - Opšta sistematizacija naučnih oblasti, Knjiga i pismo. Bibliotekarstvo. Muzeologija, Bibliotekarstvo, Specijalne biblioteke
  </dc:subject>
  ▼<dc:description>
    Bibliotekarstvo i informatika - Specijalne biblioteke Datum odbrane: 24.04.2013.
  </dc:description>
  <dc:description>Library an information science - Special libraries</dc:description>
  <dc:identifier>o:5568</dc:identifier>
  <dc:format>PDF/A (212 listova)</dc:format>
  <dc:type>Text</dc:type>
  <dc:title>Biblioteke verskih zajednica i institucija</dc:title>
  <dc:contributor>Vraneš, Aleksandra, 1960-</dc:contributor>
  <dc:contributor>Tanasković, Darko, 1948-</dc:contributor>
  <dc:contributor>Šijaković, Bogoljub, 1955-</dc:contributor>
</oai_dc:dc>

```

Слика 8 – Пример записа за исту докторску дисертацију у формату за експорт података Dublin Core XML у оквиру репозиторијума PHAIDRA

Претраживач у систему PHAIDRA не даје увид у укупан број дигиталних објеката похрањених у њему. Могуће је једино утврдити број дигиталних објеката у сваком појединачном универзитетском репозиторијуму према врсти документа. Дигитални репозиторијум Универзитета у Београду садржи 4463 документа у PDF формату, 132 слике, 12 књига, 5 видео записа, 4 електронска извора, 1 аудио запис, 12 осталих тј. некатегоризованих докумената (текстуални документи у DOC формату, презентације у PPT формату).²⁰⁶ Као PDF документ похрањено је 1.583 докторских дисертација одбрањених на Универзитету у Београду.²⁰⁷

У репозиторијум Универзитета у Крагујевцу до сада је похрањено 314 докумената у PDF формату (од чега 284 докторске дисертације) и 8 слика „а у репозиторијум Универзитета у Нишу 570 докумената у PDF формату (од којих је 161 дисертација) и 8 слика.²⁰⁸

На свим наведеним универзитетима, унос метаподатака и постављање дигиталних објеката врше углавном библиотекари универзитетских библиотека (унос су вршили и студенти Катедре за библиотекарство и информатику, као и поједини истраживачи). За сваки од универзитета омогућен је пренос података из локалних

²⁰⁶Стање на дан 20.10.2015.

²⁰⁷У Дигитални репозиторијум Универзитета у Београду, као засебна колекција, похрањују се и дигитални објекти докторских дисертација одбрањених на Универзитету у Приштини са седиштем у Косовској Митровици. До 20.10.2015. укупно је похрањено 45 докторских дисертација.

²⁰⁸Стање на дан 20.10.2015.

COBISS база, али не директно у систем PHAIDRA, већ посредством апликације која је првенствено креирана за пренос и похрањивање података о докторским дисертацијама (апликације е-тезе за сваки појединачни универзитет).²⁰⁹

Иако систем PHAIDRA нуди једноставан унос и мали број обавезних метаподатака, и после две године од завршетка Темпус пројекта још увек није пракса да наставници и сарадници универзитета сами постављају дигиталне објекте са својим радовима. Разлог за то делимично лежи у чињеници да Репозиторијум PHAIDRA не омогућава једноставну размену метаподатака између система, па остаје доста посла на поправкама и премештању метаподатака, делимично јер корисницима сучеље није конфорно и није јасно где шта треба унети, а делимично и у томе што је функција традиционалне универзитетске библиотеке да прати и бележи податке о научно-истраживачком раду на универзитету, па се и у дигиталној ери од тога још увек не одступа.

Дигитална библиотека Филолошког факултета

Имплементација

Дигитална библиотека Филолошког факултета у Београду је својеврсан дигитални институционални репозиторијум и резултат је процеса дигитализације, на првом месту, издања Филолошког факултета „у областима изучавања 34 језика и њима припадајућих књижевности и култура, као и у областима библиотекарства, информатике, архивистике и музеологије.“²¹⁰ Подршку у процесу дигитализације и креирању апликације пружила је фирма МФЦ Микрокомерц²¹¹ која је софтверски део извела уз примену Мајкрософтове технологије.

Основни постулати на којима поменута фирма развија дигиталне библиотеке у оквиру веб апликације LibIT су:

²⁰⁹ „Систем „Е-тезе“ намењен је претраживању и преузимању докторских дисертација из локалног репозиторијума, као и из узајамне базе података докторских дисертација универзитета у Београду, Нишу и Крагујевцу. Локални и узајамни репозиторијум докторских дисертација чине интегрални део дигиталних репозиторијума – PHAIDRA у којима се индексирају и архивирају научни и стручни радови наставника и сарадника Универзитета у Београду, Универзитета у Нишу и Универзитета у Крагујевцу. У складу са политиком заштите ауторских права, репозиторијуми садрже дисертације које имају потписане изјаве којима кандидат/докторанд овлашћује универзитет да у дигитални репозиторијум унесе дисертацију под одређеним лиценцим условима. Дисертације је могуће претраживати у систему „Е-тезе“ или у оквиру универзитетског репозиторијума PHAIDRA.“ – Универзитет у Београду. *Е-тезе Универзитета у Београду*, <http://eteze.bg.ac.rs/> (приступљено 30.10.2015)

²¹⁰ Александра Вранеш. „Дигитална библиотека“, <http://www.fil.bg.ac.rs/lang/sr/biblioteke/digitalna-biblioteka/> (приступљено 20.08.2014)

²¹¹ МФЦ Микрокомерц, доступно на: <http://www.earhiva.com/>

- Коришћење публикација без хабања и оштећења;
- Брз и лак приступ садржајима;
- Претрага по било којој речи, кључним речима и унапред дефинисаним атрибутима;
- Преглед дигитализованих страница;
- Чување изабраних публикација или делова публикација у PDF формату;
- Доступност садржаја заинтересованом кориснику у сваком тренутку.²¹²

Такође, захваљујући софтверу Spider.Net омогућено је и:

- Чување докумената у различитим форматима попут TIFF, JPEG, MSOfficeiOpenOffice, AutoCAD, MP3, XML;
- Класификација докумената по типовима уз доделу одговарајућих метаподатака;
- Брзо и разноврсно претраживање;
- Генерисање документа у PDF формату;
- Чување података у XML формату;
- Употреба шифрарника и сл.²¹³

Садржај

Садржај Дигиталне библиотеке Филолошког факултета чине публикације значајне за процес образовања и научног усавршавања запослениху овој установи. Публикације се налазе у отвореном приступу, а могу се претраживати и читати уз регистрацију и добијање корисничког налога.

До сада је за Дигиталну библиотеку Филолошког факултета дигитализовано преко 120.000 страна различитих публикација:

- 333 броја периодичних публикација које издаје Факултет: Анали Филолошког факултета, Прилози, Књижевност и језик, Славистика, Италика, Годишњак Катедре за српску књижевност, Филолошки преглед;
- 16 зборника радова насталих као резултат излагања на домаћим и међународним научним конференцијама у организацији Филолошког факултета
- 35 монографских публикација;

²¹² МФЦ Микрокомерс. Софтвер Spider.Net дигитална библиотека – LibIT, <http://www.earhiva.com/softver/mfc/libit.php#close> (приступљено 25.11.2014)

²¹³ МФЦ Микрокомерс- Софтвер Spider.Net репозиторијум докумената, <http://www.earhiva.com/softver/mfc/repozitorijum.php> (приступљено 25.11.2014)

- антологије, збирке;
- докторске дисертације брањене на Филолошком факултету.

У следећој фази, за потребе овог репозиторијума биће дигитализовано и рукописно књижевно наслеђе, нарочито рукописне заоставштине Иве Андрића, Десанке Максимовић и Ивана В. Лалића са чиме су се сагласили носиоци ауторских права поменутих аутора.

Метаподаци²¹⁴

Описни (дескриптивни) метаподаци укључују унос: аутора (на српском и страном језику), уредника (приређивача), преводиоца, наслова (на српском и страном језику), изавача, године издавања, међународног стандардног броја (ISBN или ISSN), кључне речи, УДК броја.

Слика 9 – Дигитална библиотека Филолошког факултета – маска за претрагу преко библиографских метаподатака

Структурални метаподаци омогућавају претрагу по монографској или серијској публикацији, као и по саставним деловима, односно поглављима и чланцима и у оквиру ових публикација.

²¹⁴Приликом описа коришћених метаподатака потребно је напоменути да креатор Дигиталне библиотеке Филолошког факултета, МФЦ Микромерц, и поред бројних молби, аутору није доставио податке о коришћеним форамтима за унос метаподатака, као ни о обавезним подацима које је потребно унети у опис дигиталног објекта. Стога смо у овом истраживању описали само оне метаподатке који су били очигледно видљиви из корисничког сучеља за претрагу.

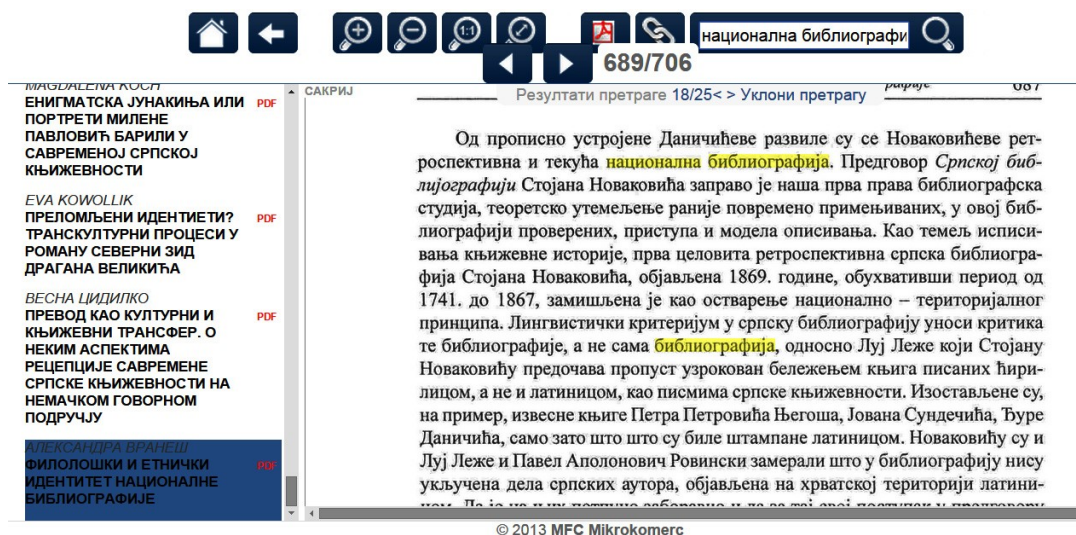
The screenshot shows a search interface with a dark blue header containing three tabs: 'ПУБЛИКАЦИЈЕ', 'ЧЛАНЦИ/ПОГЛАВЉА', and 'Списак публикација'. Below the header, there are two main search panels. The left panel, titled 'Претрага текста', contains input fields for: 'Наслов текста', 'Наслов текста на страном језику', 'Аутор текста', 'Аутор текста на страном језику', 'Име приређивача / уредника', 'Преводилац текста', and 'Кључне речи'. The right panel, titled 'Претрага по матичној публикацији', contains input fields for: 'Наслов', 'Име приређивача / уредника', 'Наслов на страном језику', 'Издавач', 'Име аутора', 'Година издавања', 'Име аутора на страном језику', 'ISBN/ISSN', 'Име преводиоца', and 'УДК број'. Below these panels is an 'Општа претрага' section with a wide search bar and a 'Тражи' button. At the bottom left, there is a small text box that says 'Попуните поља за претрагу'.

Слика 10 – Дигитална библиотека Филолошког факултета – маска за претрагу преко структуралних и библиографских метаподатака

Претрага преко комплетног текста

Једна од намена овог дигиталног репозиторијума јесте да повећа видљивост српског језика на оба писма у интернет окружењу. Софтверско решење, развијено на Мајкрософтовој технологији, предвидело је оптичко препознавање карактера, а самим тим и претраживање комплетног текста дигиталних објеката тј. преко сваке речи у тексту. Упити се могу постављати и на ћириличном и на латиничном писму, без обзира на писмо оригиналне публикације.

Из примера на слици 11 се види могућност претраге дигиталне колекције преко сваке речи у тексту, а не само преко метаподатака који су додељени конкретном дигиталном објекту. На постављени упит "национална библиографија" добијено је 25 резултата претраге. Одабрали смо 18. резултат претраге и појављивање траженог појма и у наслову рада др Александре Вранеш "Филолошки и етнички идентитет националне библиографије" и у самом тексту "... развиле су се Новаковићева ретроспективна и текућа национална библиографија." Оваква врста претраге је изузетно значајна за корисника нарочито у ситуацијама када тема рада није садржана у наслову или појам који корисник тражи није наведен у класификацији објекта (односно у метаподацима).



Слика 11 – Дигитална библиотека Филолошког факултета – пример претраге преко сваке речи у тексту

в. Еуропеана

Еуропеана је портал који је покренут 2008. године као пројекат Европске комисије са циљем да се корисницима интернета са једне приступне тачке учини доступним дигитализовано културно наслеђе европских народа. На овом порталу могуће је претраживати метаподатке о дигиталним библиотекама и дигиталним објектима и читати и прегледати њихов садржај, а преко њих и фондове најзначајнијих европских библиотека, музеја, архива и аудио-визуелних архива, па се Еуропеана још назива и Европском дигиталном библиотеком, музејем и архивом.

Путем овог портала могуће је претраживање дигитализованих текстова, сликовне грађе, видео и аудио записа, као и тродимензионалних објеката и метаподатака о њима. Текстуални објекти се односе на књиге, писма, архивску грађу, дисертације, новине, чланке и рукописе. Сливковна грађа подразумева уметничке слике, цртеже, фотографије, слике музејских експоната и нотне записе. Видео записи се односе на филмове, серије и телевизијске програме, а аудио записи на музичке и говорне секвенце са дигитализованих грамофонских плоча, трака, дискова или радио станица. Записи за тродимензионалне објекте у вези су са виртуелним 3D презентацијама, архитектонским објектима и одабраним локацијама.²¹⁵

Различитост грађе и разноврсност културних установа који учествују у пројектима Еуропеане, јасан су показатељ зашто овај портал нуди приступ до преко 30 милиона дигиталних објеката и што у поменути пројектима учествује више око

²¹⁵Europeana Portal доступно на: http://www.europeana.eu/portal/usingeuropeana_results.html

2.300 институција из 36 земаља Европе. Према подацима из новембра 2013. установе из Немачке и Француске учествују са највећим бројем дигиталних објеката (15%, односно 10,8%).²¹⁶ У оквиру појединих пројеката, корисницима портала је омогућено да сами уносе дигиталне објекте и метаподатке.²¹⁷

Треба имати у виду да Еуропеана није место или централни сервер на коме се похрањују дигитални објекти, она је приступна тачка преко које се долази до дигиталних објеката који су похрањени на серверима установа које учествују у пројектима Еуропеане. Установе су само у обавези да Еуропеани доставе метаподатке о дигиталним објектима и трајне линкове који воде ка њима, као и да промовишу њене садржаје и дигиталне колекције. Еуропеана подстиче отворени приступ и доступ документима у пуном тексту. Одлука о садржајима који ће бити понуђени корисницима Еуропеане је на појединачним установама.

Како је Еуропеана полазно место за истраживања и претраживања европског и светског културног и научног наслеђа, она је и је приступна тачка одакле је могуће претраживати и дигитализоване фондове српских установа културе и науке, у првом реду Народне библиотеке Србије и Универзитетске библиотеке “Светозар Марковић.”

Народна библиотека Србије на порталу Еуропеана

Као пуноправна чланица Конференције националних библиотека Европе (CENL²¹⁸), Народна библиотека Србије је од 1. јула 2005. године на порталу Еуропеане заступљена са узајамним каталогом COBIB.SR и метаподацима у око 3 милиона каталожских записа, а преко пројекта „Европска библиотека“ („The European Library“) чији су основни циљеви: изградња заједничког каталога европских националних библиотека који ће бити претражив на језицима свих пуноправних сарадника, изградња портала са информацијама о свим националним библиотекама Европе и изградња дигиталне европске библиотеке као трезора европске културне баштине.²¹⁹

²¹⁶ Europeana portal, <http://www.pro.europeana.eu/web/guest/content> (приступљено 20.05.2014)

²¹⁷ Реч је о тзв. User-Generated Content, тренутно важи само за пројекат Еуропеана колекције 1914-1918

²¹⁸ CENL – Conference of European National Libraries, доступно на: <http://www.cenl.org/>

²¹⁹ Весна Ињац. „Народна библиотека Србије – пуноправан партнер у пројекту 'Европска библиотека'.” *Преглед НИЦД* 7(2005):50

The screenshot shows the search interface of The European Library. At the top, the logo and tagline 'Connecting knowledge' are visible. The search bar contains the query 'seobe+milos+crnjanski'. Below the search bar, there are navigation tabs: 'HOME', 'DISCOVER', and 'ACCESS OUR DATA'. The search results are displayed in a grid format, with three items listed. Each item includes a small image of a book cover, the title 'Seobe', the library name, the author 'Miloš Crnjanski', the publication details, and the contributor. To the right of the search results, there is a 'REFINE' section with various filters like 'Accessibility', 'Discipline', 'Language', 'Contributor', and 'Collection'. Below the filters, there is a 'YEAR OF PUBLICATION' section with a bar chart showing the distribution of results by year.

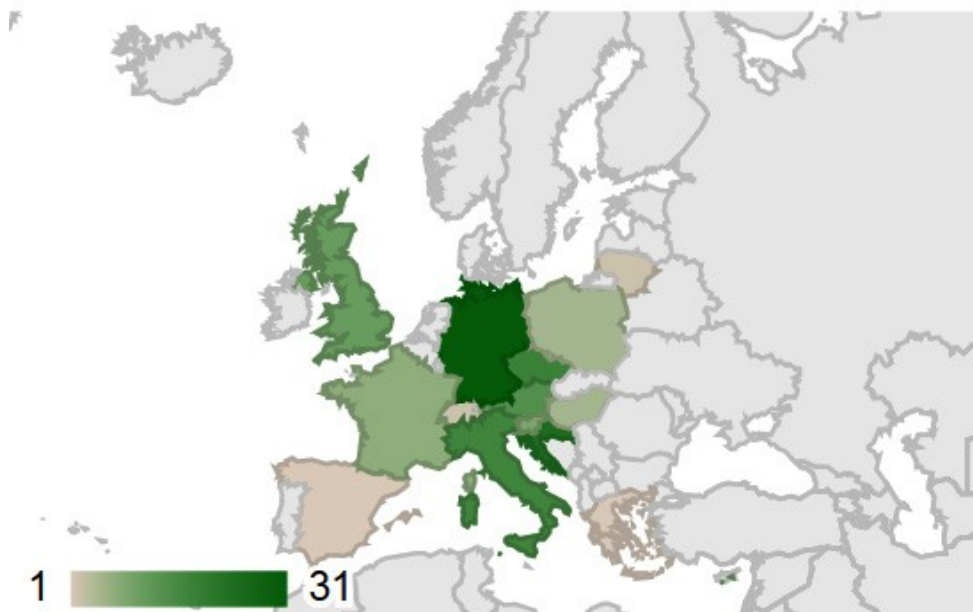
Слика 12 – Претрага на порталу „Европске библиотеке“ – испис по библиотекама у којима се публикација налази

У „Европску библиотеку“ до сада је укључено 48 националних и водећих научних библиотека Европе. Корисницима „Европске библиотеке“ данас је омогућено претраживање преко 23 милиона дигиталних објеката и преко 162 милиона библиографских записа у каталозима европских библиотека.²²⁰

Као што се може видети на слици 12, „Европска библиотека“ нуди претрагу по различитим критеријумима, а испис података је по библиотекама у којима се тражена публикација налази. Такође, постоји и графички приказ по земљама у којима су библиотеке које поседују одређену публикацију (слика 13).

²²⁰The European Library (TEL), доступно на: <http://www.theeuropeanlibrary.org/tel4/>

COUNTRY OF CONTRIBUTOR



Слика 13 – Претрага на порталу „Европске библиотеке“ – графички приказ по земљама у којима су библиотеке које поседују одређену публикацију²²¹

Поред учешћа у наведеном пројекту, Народна библиотека Србије, као једна од десет националних библиотека Европе, учествује у још једном пројекту Европеане - „Европске колекције 1914-1918.“ Овај пројекат је везан за грађу и историјске материјале из Првог светског рата. Трогодишњим пројектом руководи Државна библиотека у Берлину, а питања дигитализације, доступности комплетном тексту докумената, оптичком преознавању карактера и агрегацији садржаја на порталу Европеане поверени су Националној библиотеци Аустрије. Предвиђено је да након завршетка пројекта корисницима буду доступни метаподаци и комплетан текст 400.000 дигиталних објеката.²²² Народна библиотека Србије, у оквиру овог пројекта, уређује дигиталну библиотеку „Велики рат“ која обухвата различите дигиталне објекте – књиге, периодичку, рукописну грађу, плакате, слике, картографску грађу из поменутог

²²¹Различите боје означавају различити број примерака одређене публикације – од једне публикације у грчким до 31 публикације у немачким библиотекама.

²²² Европске колекције 1914-1918 (Europeana Collections 1914-1918), доступно на: <http://www.europeana-collections-1914-1918.eu/>

периода који дочаравају колективна или индивидуална искуства непосредних актера овог историјског догађаја.²²³

The screenshot shows the Europeana website interface. At the top left is the Europeana logo with the text '1914-1918'. Navigation links include 'Home', 'Add your story', and 'Browse'. On the top right, there are links for 'Sign in', 'Register', and a language selection dropdown. A search bar is located below the navigation. The main content area displays a search result for 'Kurrendierungsliste Nr. 18'. On the left, there is a thumbnail image of a document page. To the right of the thumbnail, the record details are listed: Description (am 21. Februar 1917 (21.02.1917)), Language (немачки (ge)), Type (leaf, official document; leaf), Subject (World War, 1914-1918 -- Serbia; World War, 1914-1918--Serbia; World War, 1914-1918 -- Austria; World War, 1914-1918--Austria; World War I; First World War; Great War; WW1; WWI; World War 1; World War I; World War One), Date (21.02.1917), Rights (Архив Србије), Copyright (© Free Access - Rights Reserved), Identifier (http://velikirat.nb.rs/items/show/3580), Data provider (National Library of Serbia), Provider (The European Library), and Providing country (serbia). Below the thumbnail, there are options to view the item at the National Library of Serbia, share on social media (Facebook, Twitter, Email, Print, Code), and a Microsoft Translator widget.

Слика 14 – Пример записа на порталу Еуропеана у оквиру пројекта „Европске колекције 1914-1918“

На сликама 14 и 15 представљени су примери записа са метаподацима за дигиталне објекте у оквиру овог пројекта (на порталу Еуропеана и у дигиталној библиотеци „Велики рат“) као и сам дигитални објекат чији је комплетан текст могуће читати и претраживати.

²²³ Народна библиотека Србије. Портал Велики рат, доступно на: <http://velikirat.nb.rs/onama>



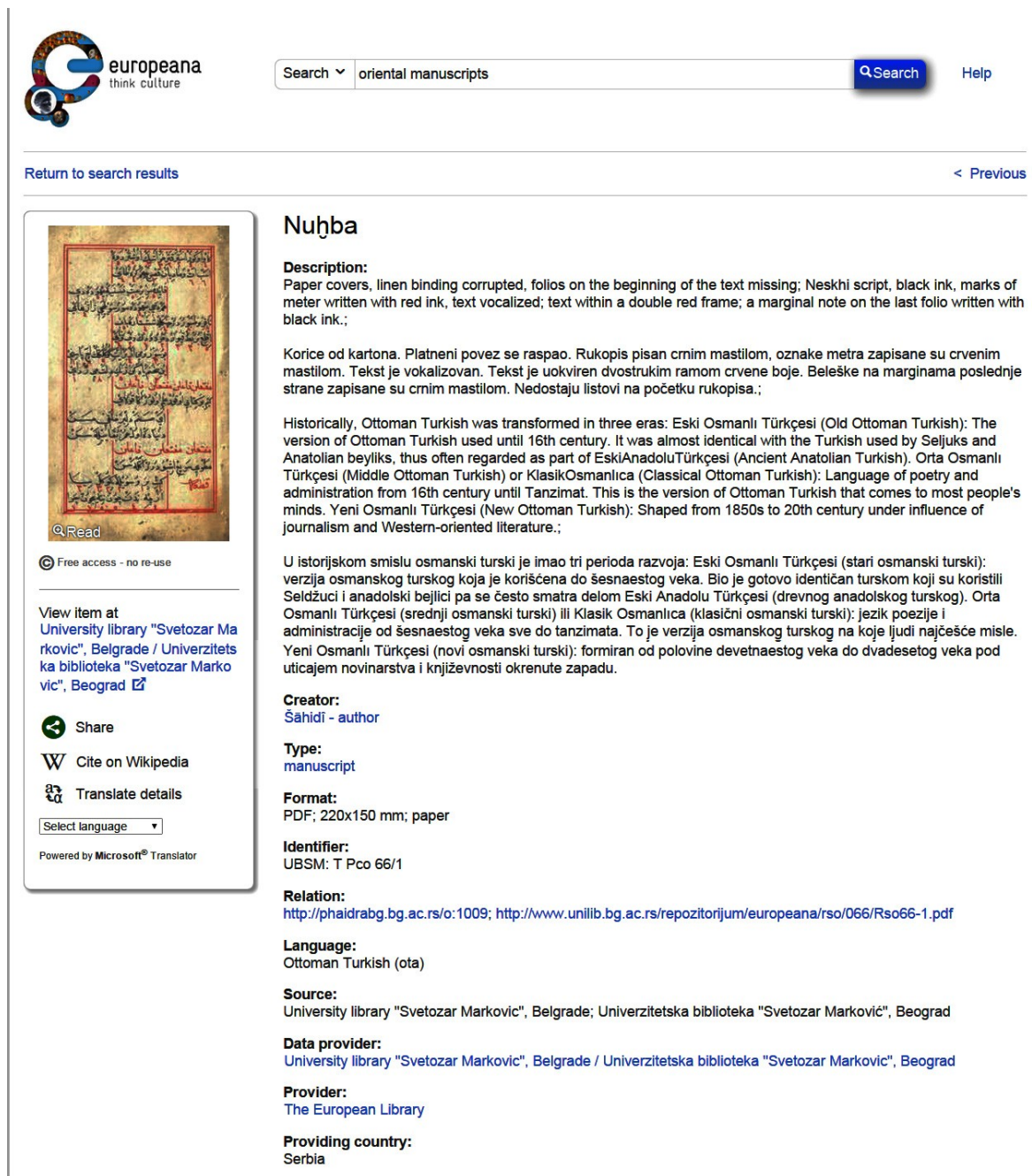
Слика 15 – Пример записа са слике 14 у дигиталној библиотеци „Велики рат“

Универзитетска библиотека „Светозар Марковић“ на порталу Еуропеана

Универзитетска библиотека „Светозар Марковић“ у Београду је учесник два Еуропеана пројекта у оквиру којих је корисницима портала дала на коришћење своје значајне фондове. У првом реду, то је пројекат “Библиотеке Еуропеане: агрегација дигиталног садржаја европских библиотека.” Пројекат је трајао од 2011. до 2013. године и у њему је учествовало 25 научних библиотека и библиотечких удружења. Овим пројектом је Универзитетска библиотека у Београду представила своје две дигиталне колекције:

1. Колекција о Александру Македонском од око 150 књига у 300 свезака (28.000 дигитализованих страна) насталих у периоду од 15. до 19. века коју је формирао Хајнрих Кристенсен (1849-1912), универзитетски професор из Хамбурга, а која се у Универзитетској библиотеци чува као посебна библиотека. Дела у овој колекцији су углавном на немачком језику, а колекција садржи и десетак докторских дисертација чија је тема Александар Македонски. Најстарија публикација је из 1494. године.
2. Дигиталну збирку 686 оријенталних рукописа из 395 томова (166.000 дигитализованих страна) писаних у периоду од 13. до 18. века. Рукописи су на арапском (525), отоманском турском (149) и персијском језику (13 рукописа). Сви рукописи потичу из европског дела Отоманског царства и писани су под јаким

утицајем локалне традиције и културе, ортографије и језика који се користио у поменутом региону. Најстарији рукопис је из 1206. године.²²⁴



The screenshot shows the Europeana search results page for the manuscript 'Nuḥba'. The interface includes the Europeana logo, a search bar with the query 'oriental manuscripts', and navigation links like 'Return to search results' and '< Previous'. The main content area displays the manuscript's title, a thumbnail image of the manuscript page, and detailed metadata. The metadata includes a description, historical context, creator information, type, format, identifier, relation, language, source, data provider, and provider details.

Europeana
think culture

Search Help

[Return to search results](#) [< Previous](#)

Nuḥba

Description:
Paper covers, linen binding corrupted, folios on the beginning of the text missing; Neskhī script, black ink, marks of meter written with red ink, text vocalized; text within a double red frame; a marginal note on the last folio written with black ink.;

Korice od kartona. Platneni povež se raspao. Rukopis pisan crnim mastilom, oznake metra zapisane su crvenim mastilom. Tekst je vokalizovan. Tekst je uokviren dvostrukim ramom crvene boje. Beleške na marginama poslednje strane zapisane su crnim mastilom. Nedostaju listovi na početku rukopisa.;

Historically, Ottoman Turkish was transformed in three eras: Eski Osmanlı Türkçesi (Old Ottoman Turkish): The version of Ottoman Turkish used until 16th century. It was almost identical with the Turkish used by Seljuks and Anatolian beyliks, thus often regarded as part of Eski Anadolu Türkçesi (Ancient Anatolian Turkish). Orta Osmanlı Türkçesi (Middle Ottoman Turkish) or Klasik Osmanlıca (Classical Ottoman Turkish): Language of poetry and administration from 16th century until Tanzimat. This is the version of Ottoman Turkish that comes to most people's minds. Yeni Osmanlı Türkçesi (New Ottoman Turkish): Shaped from 1850s to 20th century under influence of journalism and Western-oriented literature.;

U istorijskom smislu osmanski turski je imao tri perioda razvoja: Eski Osmanlı Türkçesi (stari osmanski turski): verzija osmanskog turskog koja je korišćena do šesnaestog veka. Bio je gotovo identičan turskom koji su koristili Seldžuci i anadolski bejlici pa se često smatra delom Eski Anadolu Türkçesi (drevnog anadolskog turskog). Orta Osmanlı Türkçesi (srednji osmanski turski) ili Klasik Osmanlıca (klasični osmanski turski): jezik poezije i administracije od šesnaestog veka sve do tanzimata. To je verzija osmanskog turskog na koje ljudi najčešće misle. Yeni Osmanlı Türkçesi (novi osmanski turski): formiran od polovine devetnaestog veka do dvadesetog veka pod uticajem novinarstva i književnosti okrenute zapadu.

Creator:
Şahidi - author

Type:
manuscript

Format:
PDF; 220x150 mm; paper

Identifier:
UBSM: T Pco 66/1

Relation:
<http://phaidrabbg.bg.ac.rs/o:1009>; <http://www.unilib.bg.ac.rs/repozitorijum/europeana/rso/066/Rso66-1.pdf>

Language:
Ottoman Turkish (ota)

Source:
University library "Svetozar Markovic", Belgrade; Univerzitetska biblioteka "Svetozar Marković", Beograd

Data provider:
University library "Svetozar Markovic", Belgrade / Univerzitetska biblioteka "Svetozar Markovic", Beograd

Provider:
The European Library

Providing country:
Serbia

Free access - no re-use

View item at University library "Svetozar Markovic", Belgrade / Univerzitetska biblioteka "Svetozar Markovic", Beograd

Share

Cite on Wikipedia

Translate details

Select language

Powered by Microsoft® Translator

Слика 16 – Пример записа из збирке оријенталних рукописа на порталу Европеана

На порталу Европеана могу се претраживати метаподаци за сваки дигитални објекат из ове две збирке, а пун текст у PDF формату доступан је у репозиторијуму Универзитета у Београду и репозиторијуму дигиталних збирки Универзитетске библиотеке „Светозар Марковић.“

²²⁴Adam Sofronijević and Jelena Mitrović. „Europeana Libraries Project: Fostering Digitization Across the Europe.” *NCD Review* 21(2012):62

```

▼<ns0:uwmmetadata xmlns:ns0="http://phaidra.univie.ac.at/XML/metadata/V1.0"
  xmlns:ns1="http://phaidra.univie.ac.at/XML/metadata/lom/V1.0"
  xmlns:ns10="http://phaidra.univie.ac.at/XML/metadata/provenience/V1.0"
  xmlns:ns11="http://phaidra.univie.ac.at/XML/metadata/provenience/V1.0/entity"
  xmlns:ns12="http://phaidra.univie.ac.at/XML/metadata/digitalbook/V1.0"
  xmlns:ns13="http://phaidra.univie.ac.at/XML/metadata/etheses/V1.0"
  xmlns:ns2="http://phaidra.univie.ac.at/XML/metadata/extended/V1.0"
  xmlns:ns3="http://phaidra.univie.ac.at/XML/metadata/lom/V1.0/entity"
  xmlns:ns4="http://phaidra.univie.ac.at/XML/metadata/lom/V1.0/requirement"
  xmlns:ns5="http://phaidra.univie.ac.at/XML/metadata/lom/V1.0/educational"
  xmlns:ns6="http://phaidra.univie.ac.at/XML/metadata/lom/V1.0/annotation"
  xmlns:ns7="http://phaidra.univie.ac.at/XML/metadata/lom/V1.0/classification"
  xmlns:ns8="http://phaidra.univie.ac.at/XML/metadata/lom/V1.0/organization"
  xmlns:ns9="http://phaidra.univie.ac.at/XML/metadata/histkult/V1.0">
  ▼<ns1:general>
    <ns1:identifier>o:1011</ns1:identifier>
    <ns1:title language="tr">Nuhba</ns1:title>
    <ns1:language>tr</ns1:language>
    ▼<ns1:description language="en">
      Historically, Ottoman Turkish was transformed in three eras: EskiOsmanlıTürkçesi (Old Ottoman Turkish): The version of Ottoman Turkish used until 16th century. It was almost identical with the Turkish used by Seljuks and Anatolian beyliks, thus often regarded as part of EskiAnadoluTürkçesi (Ancient Anatolian Turkish). OrtaOsmanlıTürkçesi (Middle Ottoman Turkish) or KlasikOsmanlıca (Classical Ottoman Turkish): Language of poetry and administration from 16th century until Tanzimat. This is the version of Ottoman Turkish that comes to most people's minds. YeniOsmanlıTürkçesi (New Ottoman Turkish): Shaped from 1850s to 20th century under influence of journalism and Western-oriented literature.
    </ns1:description>
    ▼<ns1:description language="sr">
      Eski Osmanlı Türkçesi (stari osmanski turski): verzija osmanskog turskog koja je korišćena do šesnaestog veka. Bio je gotovo identičan turskom koji su koristili Seldžuci i anadolski bejlici pa se često smatra delom Eski Anadolu Türkçesi (drevnog anadolskog turskog). Orta Osmanlı Türkçesi (srednji osmanski turski) ili Klasik Osmanlıca (klasični osmanski turski): jezik poezije i administracije od šesnaestog veka sve do tanzimata. To je verzija osmanskog turskog na koje ljudi najčešće misle. Yeni Osmanlı Türkçesi (novi osmanski turski): formiran od polovine devetnaestog veka do dvadesetog veka pod uticajem novinarstva i književnosti okrenute zapadu.
    </ns1:description>
    <ns2:irdata>no</ns2:irdata>
  </ns1:general>
  ▼<ns1:lifecycle>
    <ns1:upload_date>2012-03-15T12:34:38.684Z</ns1:upload_date>
    <ns1:status>44</ns1:status>
    <ns2:peer_reviewed>no</ns2:peer_reviewed>
    ▼<ns1:contribute seq="0">
      <ns1:role>46</ns1:role>
      ▼<ns1:entity seq="0">
        <ns3:firstname>Şahidi</ns3:firstname>
      </ns1:entity>
    </ns1:contribute>
  </ns1:lifecycle>
  ▼<ns1:technical>
    <ns1:format>application/pdf</ns1:format>
    <ns1:size>3457924</ns1:size>
    <ns1:location>http://phaidrabg.bg.ac.rs/o:1011</ns1:location>
  </ns1:technical>
  ▼<ns1:rights>
    <ns1:cost>no</ns1:cost>
    <ns1:copyright>yes</ns1:copyright>
    <ns1:license>2</ns1:license>
  </ns1:rights>
  ▼<ns1:classification>
    <ns1:purpose>70</ns1:purpose>
  </ns1:classification>
  ▼<ns12:digitalbook>
    <ns12:name_magazine language="tr">Nuhba</ns12:name_magazine>
  </ns12:digitalbook>
</ns0:uwmmetadata>
    
```

Слика 17 – Пример записа из збирке оријенталних рукописа у репозиторијуму PHAIDRA

Универзитетска библиотека „Светозар Марковић“ једна је од 18 пуноправних партнерских институција које су од фебруара 2012. до јануара 2015. године учествовале у пројекту „Европске новине“ (Europeana Newspapers). Пројекат је имао за циљ да корисницима Европеане омогући претрагу метаподатака и доступ до комплетног текста на 18 милиона новинских страница, као и напредну претрагу преко појединих термина, наслова чланака, именованих ентитета²²⁵, имена новинских

²²⁵Именовани ентитети „идентификују у документу властита имена и друге називе и обично их и класификују као ознаке људи, институција, догађаја, места и слично.“ – Цветана Крстев. *Материјали за курс Напредне методе у*

колекција, датума, бројева, приступ свакој појединачној новинској страни и повезивање са релевантним извором.

Пројектом је руководила Национална библиотека Холандије, а Универзитет у Инсбруку, Универзитет у Салфорду и фирма CCS – Content Conversion Specialists су радили на развоју апликација и тестирању оптичког препознавања карактера (OCR), оптичке сегментације чланака (Optical Layout Recognition – OLR) и препознавању именованих ентитета (Named Entity Recognition – NER). У оквиру овог пројекта, одабране су новине које покривају период од 1618. до 2002. године на преко 20 језика и 14 различитих формата за метаподатке.²²⁷

The screenshot shows the search interface of The European Library. At the top, there is a search bar with the text "Search all European Library content..." and a "GO" button. Below the search bar are navigation links: HOME, DISCOVER, and ACCESS OUR DATA. The main content area shows search results for "Nedelja" from 1910. It includes a list of filters (Everything, Newspapers, Collections, Full text, Remote Search) and a summary of the search criteria: "You are looking at 4 items out of 4 for advanced((TITLE, nedelja)) with contributor University of Belgrade with decades of publication 1910-1919 with month of publication March with newspaper title Nedelja with year of publication 1910 sorted by date descending". There are also options for "Items per page" (set to 10) and "Sort by" (set to Date (descending)). A preview of a newspaper page is shown with the title "Nedelja / 28 Mar 1910/ Page 1" and a short text snippet. On the right side, there is a "REFINE" section with a search bar and a "GO" button, and a list of filters with their current values: Contributor: University of B..., Decades of Publ...: 1910-1919, Month of Public...: March, Title: Nedelja, Year of Publica...: 1910, Contributor: University of Bel... (4), Newspaper Title, Decades of Publication, Year of Publication, and Month of Publication.

Слика 18 – Пример записа из збирке српских новина на порталу „Европска библиотека“

Универзитетска библиотека „Светозар Марковић“ је у пројекту учествовала са 45 наслова старих српских листова и новина од 1830. до 1944. године на око 400.000 дигитализованих страна. Неки од наслова су: *Време, Правда, Недеља, Звезда, Мале новине, Београдске општинске новине, Стражилово, Нова искра, Шумадинка,*

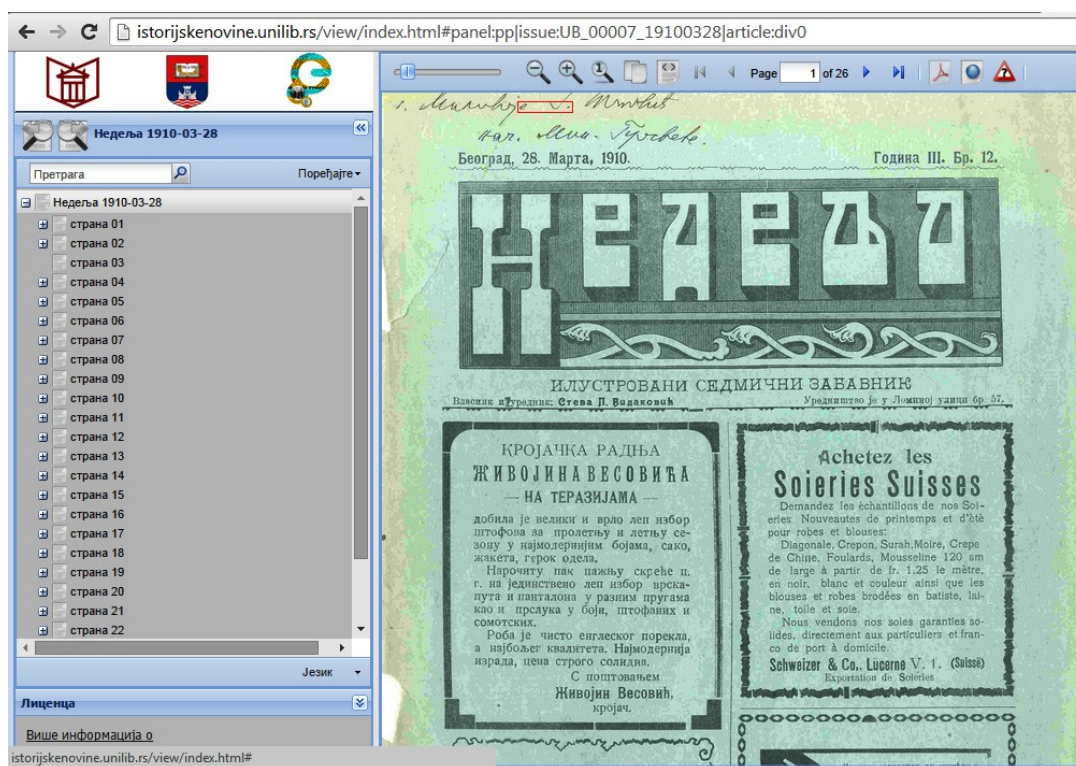
pronalaženju informacija, nepublikovani materijal, ,

<http://poincare.matf.bg.ac.rs/~cvetana/Nastava/Materijal/NM-PI-03.ppt> (приступљено 20.10.2015)

²²⁶ „Imenovani entiteti (eng. Named Entities, skr. NE) su definisani kao vlastita imena i određeni izrazi za iznose.“ – Miloš Utvić. „Konačni automati u regularnoj imenskoj derivaciji“ (Magistarska teza, Matematički fakultet, 2008), 4.

²²⁷ Clemens Neudecker and Lotte Wilms. Europeana Newspapers – a Gateway to Europeana Newspapers Online. In: *IFLA Newspapers/GENLOC Pre-Conference Satellite Meeting Newspapers to the People*, Singapore, 14-15 August 2013, http://www.ifla.org/files/assets/newspapers/Singapore_2013_papers/day_1_10_2013_ifla_satellite_neudecker_clemens_wilms_lotte_europeana_newspapers_-_a_gateway_to_european_newspapers_online.pdf (приступљено 26.08.2014)

Подунавка, Зора, Стармали, Стармлади, Бодљикаво прасе, Балкански рат у слици и речи, Илустрована ратна кроника, Ратни дневник (који је и део дигиталне библиотеке Велики рат). За потребе претраживања ове велике колекције Универзитетска библиотека „Светозар Марковић“ у сарадњи са партнером на пројекту, Националном библиотеком Луксембурга²²⁸, развила је претраживач заснован на софтверу *bnlviewer*²²⁹ (слика 19).



Слика 19 – Пример претраге збирке српских новина на претраживачу Универзитетске библиотеке „Светозар Марковић“

Метаподаци

Институције – учеснице у пројектима порталу Европеана достављају метаподатке у различитим форматима. Зависно од врсте институције (музеј, библиотека, архив, аудио-визуелни архив) и садржаји и формати се међусобно разликују. Да би се превладале разлике које постоје међу различитим форматима, препоручено је да минимални скуп метаподатака који прихвата Европеана буде Даблинско језгро, а подаци се могу доставити и у формату MARC 21. Подразумевана структура за оба формата је XML. Српске библиотеке углавном користе извоз метаподатака из система COBISS, управо у форматима DublinCore.xml или

²²⁸ Bibliothèque nationale de Luxembourg, доступно на: <http://www.bnl.public.lu/fr/index.html>

²²⁹ Bibliothèque nationale de Luxembourg. *Bnlviewer*, доступно на: <http://sourceforge.net/projects/bnlviewer/>

MARC21.xml. Како су то минимални скупови метаподатака, а да би се побољшали резултати у проналажењу информација, у оквиру пројеката Европеанесу развијени или дорађени разни системи за метаподатке, као и разне методе за агрегацију и пречишћавање текста.

Ако посматрамо формате за метаподатке, уочићемо да су креатори Европеане развили Европеана семантичке елементе ESE (Europeana Semantic Elements)²³⁰ и Модел података Европеане EDM (Europeana Data Model).²³¹ За техничке метаподатке у оквиру Европеане значајна је дорада метаподатака у моделу METS/ALTO (Metadata Encoding and Transmission Standard / Analyzed Layout and Text Object).

ESE је скуп метаподатака који се користи у опису дигиталних објеката културне баштине садржане на порталу Европеана. То је формат базиран на стандарду Даблинско језгро, а садржи и додатне елементе који помажу бољем претраживању садржаја Европеане. ESE комбинује описне и административне метаподатке тако што користи:

- 15 основних елемената Даблинског језгра (наведени у поглављу 4.2.3);
- Core Metadata Initiative Terms Specification – DC Terms²³²;

<i>dcterms:alternative</i> – алтернативни наслов	податак о додатном наслову или имену дигиталног објекта, односно ресурса; у корелацији је са dc:title
<i>dcterms:created</i> – креиран	датум креирања дигиталног објекта, односно ресурса; у корелацији је са dc:date
<i>dcterms:issue</i> – издат	формални датум издавања аналогне публикације или дигиталног објекта; у корелацији је са dc:date
<i>dcterms:spatial</i> – просторна покривеност	простор на који се односи садржај дигиталног објекта, у корелацији је са dc:coverage
<i>dcterms:tempora</i> – временска покривеност	временски период на који се односи садржај дигиталног објекта, у корелацији је са dc:coverage
<i>dcterms:extent</i> – величина	односи се на величину за текстуалне или сликовне дигиталне објекте или дужину трајања за дигиталне објекте који садрже аудио и видео записе, у корелацији је са dc:format
<i>dcterms:medium</i> – медиј	материјал или физички носилац описаног извора; у корелацији је са dc:format
<i>dcterms:provenance</i> – порекло	опис свих промена која су настале на ресурсу

²³⁰ESE – Europeana Semantic Elements, доступно на: <http://pro.europeana.eu/ese-documentation/>

²³¹ EDM– Europeana Data Model, доступно на: <http://pro.europeana.eu/edm-documentation>

²³²Dublin Core Metadata Initiative. „DCMI Metadata Terms”, <http://dublincore.org/documents/dcmi-type-vocabulary/index.shtml> (приступљено 21.10.2015)

<i>dcterms:isPartOf – је део</i>	податак о ресурсу у који је описивани ресурс укључен на физички или логички начин, у корелацији је са dc:relation
<i>dcterms:hasPart – постоји део</i>	податак о ресурсу који је укључен у описивани ресурс на физички или логички начин, у корелацији је са dc:relation
<i>dcterms:isFormatOf – је формат</i>	податак о ресурсу који је суштински исти као описивани ресурс, само је у другом формату; у корелацији је са dc:relation
<i>dcterms:hasFormat – постоји формат</i>	податак о описиваном ресурсу који је суштински исти као неки раније описан ресурс, само је у другом формату; у корелацији је са dc:relation
<i>dcterms:isVersionOf – је верзија</i>	податак о сличном ресурсу у односу на који је описивани ресурс верзија, ново издање или адаптација; у корелацији је са dc:relation
<i>dcterms:hasVersion – постоји верзија</i>	податак о сличном ресурсу који је у односу на описивани ресурс верзија, ново издање или адаптација; у корелацији је са dc:relation
<i>dcterms:isReferencedBy – реферише се</i>	податак о сличном ресурсу који реферише, цитира, или на други начин указује на описивани ресурс; у корелацији је са dc:relation
<i>dcterms:References – упућује</i>	податак о сличном ресурсу који се реферише, цитира или се на њега на други начин указује у описиваном ресурсу; у корелацији је са dc:relation
<i>dcterms:isReplacedBy – замењује се</i>	податак о сличном ресурсу који замењује описивани ресурс; у корелацији је са dc:relation
<i>dcterms:Replaces – замењује</i>	податак о сличном ресурсу који је замењен описиваним ресурсом; у корелацији је са dc:relation
<i>dcterms:isRequiredBy – захтева се</i>	податак о сличном ресурсу који захтева описивани ресурс у смислу подршке за извршење неке своје функције; у корелацији је са dc:relation
<i>dcterms:Requires – захтева</i>	податак о сличном ресурсу кога захтева описивани ресурс у смислу подршке за извршење неке своје функције; у корелацији је са dc:relation
<i>dcterms:conformsTo – у складу са</i>	податак о утврђеном стандарду са којим је описивани ресурс у складу, односно на основу ког је описивани ресурс настао; у корелацији је са dc:relation
<i>dcterms:tableOfContents – садржај</i>	подаци о јединицама које чине описивани ресурс тј. подаци о садржају описиваног ресурса; у корелацији је са dc:description

- 13 елемената које је Еуропеана дефинисала према својим потребама²³³:

²³³Europeana. „Europeana Semantic Elements Specification and Guidelines“, <http://pro.europeana.eu/documents/900548/2eee7beb-b9d8-4532-a089-8e8d6df38ce7> (приступљено 28.08.2014)

<i>europaana:isShownAt – је приказан на</i>	податак о јединственој URL адреси добављача података на којој се налази пуна информација о ресурсу; у корелацији је са dc:relation објекта, односно ресурса; у корелацији је са dc:title
<i>europaana:isShownBy – се приказује помоћу</i>	податак о јединственој URL адреси добављача података на којој се налази ресурс у најбољој резолуцији, односно квалитету; у корелацији је са dc:relation; у корелацији је са dc:date
<i>europaana:object – објекат</i>	URL адреса одговарајућег извора са сликом у најбољој могућој резолуцији која је доступна на страници добављача података, а која се у мањој резолуцији може искористити за портал Европеане; то је често иста URL адреса као и у етикети europaana:isShownBy
<i>europaana:provider – добављач</i>	име организације која доставља податке Европеани
<i>europaana:dataProvider – добављач података</i>	име или идентификатор организације која даје податке Европеани
<i>europaana:type – врста</i>	податак о врсти ресурса; односи се на термин који најбоље описује опште категорије и функције ресурса; додела термина се врши избором из контролисаног речника
<i>europaana:right – права</i>	податак о ауторским правима која се примењују у ресурсу и над њим
<i>europaana:unstored – неускладиштен</i>	скуп података који обухвата све релевантне информације које иначе не могу бити мапиране другим елементима у ESE формату
<i>europaana:country – држава</i>	податак о имену државе која доставља податке или само име „Европа” у случају ширих европских пројеката
<i>europaana:language – језик</i>	податак о језику ресурса
<i>europaana:uri – јединствени идентификатор извора</i>	податак о јединственом URI у контексту Европеане
<i>europaana:usertag – ознака корисника</i>	ознака коју је креирао корисник у контексту сучеља Европеана
<i>europaana:year – година</i>	временски податак који се односи на аналогни или изворно дигитални објекат увек у форми четвороцифреног броја који означава годину према Грегоријанском календару

Установе су дужне да обавезно попуне само 8 елемената у оквиру овог скупа:

1. dc:title или dc:description;
2. dc:language за објекте који садрже текст;
3. europaana:dataProvider;
4. europaana:isShownAt или europaana:isShownBy;
5. europaana:provider;
6. dc:subject или dc:type или dc:coverage или dcterms:spatial;
7. europaana:rights;

8. europeana:type.

Радна група Европеане те елементе допуњава са још 5 обавезних метаподатака:

1. europeana:country;
2. europeana:language;
3. europeana:uri,
4. europeana:usertag;
5. europeana:year.

Међутим, због квалитетније претраге и бољег одзива у проналажењу информација, препорука је да установе попуне што више поља са метаподацима да би сви аспекти везани за један дигитални објекат били покривени.

Да бисмо илустровали поменути формат, навешћемо пример записа у ESE формату за публикацију о Александру Великом из фонда Универзитетске библиотеке „Светозар Марковић.”

<record>

<dc:title xml:lang="ger">Studien über das Bildniss Alexanders des Grossen : ein Beitrag zur alexandrinischen Kunstgeschichte mit einem Anhang über die Anfänge des Alexander kultes : mit 13 Tafeln und 36 Taxtabbildungen</dc:title>

<dc:creator>Schreiber, Theodor (1848-1912)– author</dc:creator>

<dc:description xml:lang="eng">A study of images of Alexander the Great in the ancient Greek, Hellenistic and Roman art, written by Theodor Schreiber (1848–1912), a German classical archeologist, art historian and university professor. Schreiber published widely on Hellenistic art topics and his works weren't based only on the archeological finds, but also on the philological sources. The book is illustrated. </dc:description>

<dc:description xml:lang="srp">Studija o slikama o Aleksandru Velikom u drevnoj grčkoj, helenističkoj i rimskoj umetnosti, koju je napisao nemački arheolog, istoričar umetnosti i profesor na univerzitetu Teodor Šrajber (1848-1912). Šrajber je objavljivao o helenističkoj umetnosti i njegova dela se nisu zasnivala samo na arheološkim otkrićima već i na filološkim izvorima. Knjiga je ilustrovana.</dc:description>

<dc:language>ger</dc:language>

<dc:format>PDF</dc:format>

<dc:format xml:lang="eng">Publication has 322 pages and 13 papers with tables.</dc:format>

<dcterms:extent>28cm</dcterms:extent>

<dcterms:medium>paper</dcterms:medium>

<dc:source xml:lang="eng">University library "Svetozar Markovic", Belgrade</dc:source>

<dc:source xml:lang="srp">Univerzitetaska biblioteka "Svetozar Marković", Beograd</dc:source>

<dc:rights xml:lang="eng">University library "SvetozarMarkovic", Belgrade</dc:rights>

<dc:rights><http://www.unilib.bg.ac.rs/></dc:rights>

<dc:identifier>UBSM: ПБ4 4</dc:identifier>

<dc:publisher xml:lang="ger">Leipzig : bei B. G. Teubner</dc:publisher>

<dc:date>1903</dc:date>

<dcterms:provenance>Ekslibris Dr Heinrich Christensen</dcterms:provenance>

<dc:subject xml:lang="srp">Aleksandar Veliki (356BC-323BC)– u likovnoj umetnosti</dc:subject>

<dc:subject xml:lang="eng">Alexander the Great (356BC-323BC)– in fine arts</dc:subject>

<dc:type>monograph</dc:type>

```

<dcterms:hasFormat>http://www.unilib.bg.ac.rs/repozitorijum/europeana/pb4/004/004.pdf</dcterms:hasFormat>
<dcterms:isPartOf xml:lang="ger">Abhandlungen der Philologisch-Historischen Klasse der Königl. Sächsischen Gesellschaft der Wissenschaften ; Bd. 21, No. 3</dcterms:isPartOf>
<europeana:object>http://www.unilib.bg.ac.rs/repozitorijum/europeana/pb4/004/004.jpg</europeana:object>
<europeana:provider>The European Library</europeana:provider>
<europeana:dataProvider xml:lang="eng">University library "Svetozar Markovic", Belgrade</europeana:dataProvider>
<europeana:dataProvider xml:lang="srp">Univerzitetska biblioteka "Svetozar Marković", Beograd</europeana:dataProvider>
<europeana:type>TEXT</europeana:type>
<europeana:isShownBy>http://phaidrabg.bg.ac.rs/o:559</europeana:isShownBy>
<europeana:rights>http://www.europeana.eu/rights/rr-f</europeana:rights>
</record>

```

Као што се и из примера види, ESE садржи разноврсне метаподатке, међутим, он је и даље линеарни модел метаподатака који не омогућава повезивање дигиталног објекта са другим изворима на мрежи. Како је број дигиталних објеката у оквиру Европеане брзо растао и како ти дигитални објекти носе најразноврсније садржаје, приступило се развоју новог модела за метаподатке EDM који је заснован на технологијама и особинама семантичког веба чиме омогућава још боље проналажење информација, већу видљивост на мрежи и бољу доступност садржаја.

У документацији која прати модел EDM нови квалитет је описан на следећи начин – провајдер А је послао дигитални објекат; провајдер Б може да допуни дигитални објекат новим метаподацима, као и метаподацима из нормативне датотеке коју одржава провајдер Ц или метаподацима из мрежног тезауруса који припада провајдеру Д; порекло свих података је јасно означено, а и крајњи корисник може да учествује у допуни метаподатака.²³⁴

EDM нуди доделу метаподатака и за целине и за саставне делове публикација – како за целу дигиталну књигу, тако и за поједина поглавља, уводне текстове, регистре у оквиру ње. С друге стране, архивска документа – писма, посвете, фотографије, осим што се посматрају као одвојени дигитални објекти, путем модела EDM могу се организовати и у посебне збирке. Изузетно је важно што модел одвојено посматра дигитални објекат и скуп извора о објекту којима се приступа преко мреже. Једном дигиталном објекту може бити придружено мноштво семантичких података, као одвојених делова семантичког веба.

²³⁴Europeana. „Europeana Data Model Primer“, <http://pro.europeana.eu/documents/900548/770bdb58-c60e-4beb-a687-874639312ba5> (приступљено 28.08.2014)

Да би ефикасно функционисао према постављеним параметрима, модел EDM садржи следеће скупове елемената:

1. елементе преузете из других извора

- именски простор²³⁵ RDF стандарда и RDF Shema (RDFS)²³⁶ са елементом `rdfs:resource`;
- именски простор Иницијативе отвореног архива за поновно коришћење објеката и њихову размену – Open Archive Initiative Object Reuse and Exchange (OAI ORE)²³⁷ са следећим елементима:
`ore:aggregation`
`ore:proxy`
`ore:aggregates`
`ore:proxyFor`
`ore:proxyIn`;
- именски простор Simple Knowledge Organization System (SKOS)²³⁸ са елементом `skos:Concept`;
- именски простор Даблинског језгра у домену особина елемената, услова и врсте извора са следећим елементима:

<code>dc:contributor</code>	<code>dcterms:hasFormat</code>
<code>dc:coverage</code>	<code>dcterms:hasPart</code>
<code>dc:date</code>	<code>dcterms:hasVersion</code>
<code>dc:description</code>	<code>dcterms:isFormatOf</code>
<code>dc:format</code>	<code>dcterms:isPartOf</code>
<code>dc:identifier</code>	<code>dcterms:isReferencedBy</code>
<code>dc:language</code>	<code>dcterms:isReplacedBy</code>
<code>dc:publisher</code>	<code>dcterms:isRequiredBy</code>
<code>dc:relation</code>	<code>dcterms:issued</code>

²³⁵ Под именским простором подразумева се окружење које садржи логички груписане идентификаторе или симболе (имена), а идентификатор дефинисан у једном именском простору је у асоцијативној вези са тим именским простором иако под истим именом може постојати у другом именском простору, са истим, сличним или различитим значењем – према дефиницији појма `namespace`, доступно на: <http://www.webopedia.com/TERM/N/namespace.html>

²³⁶ RDF Shema, доступно на: http://www.w3.org/2000/01/rdf-schema#

²³⁷ Open Archive Initiative Object Reuse and Exchange, доступно на: <http://www.openarchives.org/ore/1.0/toc.html>

²³⁸ Simple Knowledge Organization System, доступно на: <http://www.w3.org/2009/08/skos-reference/skos.html#>

dc:rights	dcterms:isVersionOf
dc:source	dcterms:medium
dc:subject	dcterms:references
dc:title	dcterms:replaces
dc:type	dcterms:requires
dcterms:alternative	dcterms:spatial
dcterms:conformsTo	dcterms:tableOfContent
dcterms:created	dcterms:temporal
dcterms:extent	

- именски простор W3C Data Catalog Vocabulary (DCAT)²³⁹ са елементом dcat:Dataset.

2. елементе који су креирани за потребе модела EDM²⁴⁰ дати су у табели 1:

Табела 1 – Елементи модела EDM

EDM атрибут	Опис атрибута
<i>edm:agent</i>	појединац или група људи који су одговорни за одређену активност
<i>edm:EuropeanaAggregation</i>	скуп ресурса везаних за један објекат културне баштине који тај објекат представљају у Европеани
<i>edm:EuropeanaObject</i>	сваки објекат који је резултат активности у Европеани
<i>edm:Event</i>	догађај као скуп кохерентних друштвених појава или културних манифестација ограничених временски и просторно
<i>edm:InformationResource</i>	информациони извор – текст књиге, дигитални објекат, музички запис итд.
<i>edm:NonInformationResource</i>	сва средства која нису информациони извори – људи, места, физичке јединице итд.
<i>edm:PhysicalThing</i>	физичка јединица нпр. слика, књига, скулптура, зграда, али не и особе; углавном се односи на објекте културне баштине
<i>edm:Place</i>	место на површини земљине кугле
<i>edm:ProvidedCHO (Cultural Heritage Object)</i>	предмети културне баштине за које Европеана прикупља описе и друге информације
<i>edm:TimeSpan</i>	класа апстрактних временских опсега који имају свој почетак, трајање и крај нпр. Пелопонески рат, изградња Кипа слободе и сл.
<i>edm:WebResource</i>	информациони извор који има најмање једну веб презентацију и један URI

²³⁹W3C Data Catalog Vocabulary, доступно на: <http://www.w3.org/TR/vocab-dcat/>

²⁴⁰ Europeana. „Definition of the Europeana Data Model v5.2.5”, <http://pro.europeana.eu/documents/900548/0d0f6ec3-1905-4c4f-96c8-1d817c03123c> (приступљено 28.08.2014)

Проналажење информација у дигиталним библиотекама

<i>edm:AgreggatedCHO</i>	објекат културне баштине је део скупа збирног ресурса за који је извршена аграгација података
<i>edm:begin</i>	датум који означава почетак одређеног временског периода – нпр. датум рођења особе која се описује моделом
<i>edm:CollectionName</i>	име колекције које је додељено одређеном скупу објеката у Европеани
<i>edm:datasetName</i>	идентификатор који је додељен скупу података у Европеани
<i>edm:country</i>	име државе која доставља податке или само име „Европа” у случају ширих европских пројеката
<i>edm:currentLocation</i>	географска локација или назив места или зграде који су у вези са ресурсом
<i>edm:dataProvider</i>	име или идентификатор организације која индиректно доставља податке за агрегацију
<i>edm:end</i>	датум који означава крај одређеног временског периода
<i>edm:EuropeanaProxy</i>	податак за интерну употребу, означава проху сервер Европеане
<i>edm:happenedAt</i>	веза ка месту на коме се неки догађај одиграо
<i>edm:hasMet</i>	односи се на ресурсе са објектима или феноменима који су се одиграли истовремено или заједно са посматраним ресурсом
<i>edm:hasType</i>	веза назива ресурса са појмом у неком од тезауруса или веза нпр. слике уметничког дела са описом тог дела
<i>edm:hasView</i>	скуп података о објекту културне баштине који осим описа објекта садржи и слику објекта, апстракт и садржај
<i>edm:incorporates</i>	опис једног ресурса даје додатну вредност другом ресурсу нпр. опис једне песме из антологије са везом ка опису антологије или опис перформанса из једне позоришне представе са везом ка опису представе
<i>edm:isAnnotationOf</i>	анотација која се односи на објекат у Европеани са везом ка ресурсу који је аотиран
<i>edm:isDerivateOf</i>	дораде, смањења, проширења одређених извора, односно различите верзије једног извора
<i>edm:isNextInSequence</i>	се на два дела истог ресурса који долазе непосредно један иза другог
<i>edm:isRelatedTo</i>	контекстуалне особине које се односе на појаве које су се десиле истовремено као и ресурс који се разматра или на шта се ресурс односи у погледу облика или неке друге карактеристике
<i>edm:isRepresentationOf</i>	веза ресурса ка другом ресурсу који га представља
<i>edm:isShownAt</i>	податак о јединственој URL адреси добављача података на којој се налази пуна информација о ресурсу
<i>edm:isShownBy</i>	податак о јединственој URL адреси добављача података на којој се налази ресурс у најбољој резолуцији, односно квалитету
<i>edm:isSimilarTo</i>	део садржаја једног ресурса је у вези са идејама, обликом, структуром, бојом, темом неког другог ресурса

<i>edm:isSuccessorOf</i>	однос између наставка ресурса и самог ресурса, садржај наставака није идентичан или сличан у односу на посматрани ресурс; реч је о деловима заједничке целине нпр. трилогије, часописи и сл.
<i>edm:landingPage</i>	однос између представљања објекта културне баштине на Европеани и веб презентације тог објекта
<i>edm:language</i>	податак о језику ресурса
<i>edm:object</i>	URL адреса одговарајућег извора са сликом у најбољој могућој резолуцији која је доступна на страници добављача података, а која се у мањој резолуцији може искористити за портал Европеане
<i>edm:occurredAt</i>	догађаја са најкраћим временским периодом који се односи на тај догађај
<i>edm:preview</i>	URL адреса слике која представља дигитални објекат, а коју генерише Европеана
<i>edm:provider</i>	име организације која доставља податке Европеани
<i>edm:realizes</i>	између физичке јединице и информационог извора који је у њој садржан
<i>edm:rights</i>	податак о ауторским правима која се примењују у ресурсу и над њим
<i>edm:type</i>	податак о врсти ресурса; односи се на термин који најбоље описује опште категорије и функције ресурса; додела термина се врши избором из контролисаног речника
<i>edm:ugc</i> (user generated content)	садржај који је генерисао корисник
<i>edm:unstored</i>	скуп података који обухвата све релевантне информације које иначе не могу мапирати у друге елементе у EDM
<i>edm:uri</i>	податак о јединственом URI у контексту Европеане
<i>edm:userTag</i>	ознака коју је креирао корисник у контексту сучеља Европеана);
<i>edm:wasPresentAt</i>	повезивање људи, ствари или информационих извора са догађајима којима су присуствовали или чији су део
<i>edm:year</i>	временски податак који се односи на аналогни или изворно дигитални објекат увек у форми четвороцифреног броја који означава годину према Грегоријанском календару

У одељку 5.3.1 биће показано да овако дефинисан модел за метаподатке омогућава квалитетнији опис дигиталних објеката, његово прецизније проналажење преко више параметара за претрагу и даје могућност да се дигитални објекат повеже и са другим изворима на мрежи и тиме постане део семантичког веба.

Ради илустрације употребе модела EDM преузели смо пример са портала Европеана. Реч је о опису дигиталног објекта који представља слику тромбона који се користи у војним оркестрима (Виссип тромбон). У EDM опис уграђени су додатни

подаци, односно линкови ка додатним информацијама које су у вези са дигиталним објектом, а то су видео снимак извођења композиције на овом инструменту и аудио запис са звуком овог специфичног тромбона. Већ на овом примеру, а на основу веза ка различитим веб локацијама са мултимедијалним садржајима, видљиве су предности овог модела за метаподатке. Предности се у првом реду огледају у обједињавању свих корисних информација о дигиталном објекту на једном месту.²⁴¹

```
<edm:ProvidedCHO rdf:about="#UEDIN:214">
  <dc:identifier>#UEDIN:214</dc:identifier>
  <dc:coverage/>
  <dc:date>Circa 1840</dc:date>
  <dc:description>Technical description: Brass; ligature fitting on bell section at joint; stockings on main slides. Bell with one coil, angled to face forwards. Repair History: Main slide possibly not original (tenon of slide section of joint is tapered, bell section joint for cylindrical tenon).</dc:description>
  <dc:title>Buccin trombone. Nominal pitch: B?.</dc:title>
  <dc:type rdf:resource="http://www.mimo-db.eu/InstrumentsKeywords/4378"/>
  <dc:type rdf:resource="http://www.mimo-db.eu/HornbostelAndSachs/356"/>
  <edm:type>IMAGE</edm:type>
</edm:ProvidedCHO>
<edm:WebResource rdf:about="http://www.mimo-db.eu/media/UEDIN/VIDEO/0032195v.mpg">
  <edm:rights rdf:resource="http://creativecommons.org/licenses/by-nc-sa/3.0/">
</edm:WebResource>
<edm:WebResource rdf:about="http://www.mimo-db.eu/media/UEDIN/AUDIO/0032195s.mp3">
  <edm:rights rdf:resource="http://creativecommons.org/licenses/by-nc-sa/3.0/">
</edm:WebResource>
<edm:WebResource rdf:about="http://www.mimo-db.eu/media/UEDIN/IMAGE/0032195c.jpg">
  <edm:rights rdf:resource="http://creativecommons.org/licenses/by-nc-sa/3.0/">
</edm:WebResource>
<edm:Place rdf:about="http://sws.geonames.org/3017382/">
  <skos:preflabel xml:lang="en">France</skos:PrefLabel>
</edm:Place>
<skos:Concept rdf:about="http://www.mimo-db.eu/InstrumentsKeywords/4378">
  <skos:prefLabel xml:lang="en">Buccin</skos:prefLabel>
</skos:Concept>
<skos:Concept rdf:about="http://www.mimo-db.eu/HornbostelAndSachs/356">
  <skos:prefLabel xml:lang="en">423.22 Labrosones with slides</skos:prefLabel>
</skos:Concept>
<ore:Aggregation rdf:about="http://www.mimo-db.eu/UEDIN/214">
  <edm:aggregatedCHO rdf:resource="#UEDIN:214"/>
  <edm:hasView rdf:resource="http://www.mimo-db.eu/media/UEDIN/VIDEO/0032195v.mpg"/>
  <edm:hasView rdf:resource="http://www.mimo-db.eu/media/UEDIN/AUDIO/0032195s.mp3"/>
  <edm:hasView rdf:resource="http://www.mimo-db.eu/media/UEDIN/IMAGE/0032195c.jpg"/>
  <edm:dataProvider>University of Edinburgh</edm:dataProvider>
  <edm:provider>MIMO - Musical Instrument Museums Online</edm:provider>
  <edm:rights rdf:resource="http://creativecommons.org/licenses/by-nc-sa/3.0/">
  <edm:isShownBy rdf:resource="http://www.mimo-db.eu/media/UEDIN/IMAGE/0032195c.jpg"/>
  <edm:isShownAt rdf:resource="http://www.mimo-db.eu/UEDIN/214"/>
  <edm:object rdf:resource="http://www.mimo-db.eu/media/UEDIN/IMAGE/0032195c.jpg"/>
</ore:Aggregation>
```

²⁴¹ Пример доступан на: http://pro.europeana.eu/c/document_library/get_file?uuid=99ce6a74-8e55-4321-917a-65bdf1fe5bc&groupId=51031

Портал Европеана и семантички веб

Управо је повезивање дигиталних објеката и информација о њима путем семантичког веба један од основних задатака Европеане. Није довољно да се метаподаци о дигиталним објектима могу претраживати на једном месту већ је потребно да се успоставе семантичке везе ка другим изворима на мрежи који су на неки начин повезани са конкретним дигиталним објектом. Ова функционалност постиже се додавањем локација других мрежних извора и података и заснована је на сложенијим и проширивим моделима метаподака који се развијају у оквиру радних група Европеане.

Еuropeана своје активности базира на могућностима семантичког веба који је, према Тиму Бернерс-Лију наставак, односно проширење постојећег веба где је информацији дато прецизно дефинисано значење, а чију архитектуру чине следеће технологије:

- XML²⁴² – општа синтакса за означавање података помоћу одговарајућих етикета (ознака или тагова)²⁴³
- XML Schema²⁴⁴ – начин на који се дефинише структура XML документа и одређује да ли је неки документ усклађен са прописаном структуром; XML Schema представља скуп правила по којима се одређени елементи могу користити и дефинише односе међу елементима; XML Schema је и сама посебан XML у стандардном формату који прописује W3C;
- RDF²⁴⁵ – модел података за опис дигиталних објеката и веза између њих који се најчешће изражава XML синтаксом; RDF дефинише стандардизовани начин навођења метаподатака;
- RDF shema (RDFS)²⁴⁶ – речник помоћу кога се дефинишу особине дигиталних објеката и класификују дигитални објекти који су описани RDF моделом (термини се групишу у одређене класе);
- OWL²⁴⁷ – језик за описивање особина и класа, као и односа међу класама тј. језик за описивање онтологија²⁴⁸ у оквиру семантичког веба односно концепата

²⁴²XML – Extensible Markup Language, доступно на: <http://www.w3.org/XML/>

²⁴³У одељку 4.4.2 опширније су описани XML и RDF

²⁴⁴XML Schema, доступно на: <http://www.w3.org/standards/xml/schema>

²⁴⁵RDF – Resource Description Framework, доступно на: <http://www.w3.org/RDF/>

²⁴⁶RDF Schema, доступно на: <http://www.w3.org/TR/rdf-schema/>

и веза које постоје у оквиру система појмова; OWL садржи речник термина за метаподатке и представља проширење RDF.

Следећи овакву структуру, а због потребе за повезивањем података на вебу у семантичком, а не само у линераном смислу, креатори Еуропеане развили су описане моделе метаподатака. Као што се види из анализираних модела метаподатака, EDM користи метаподатке који су описани стандардом ESE, али као њихову семантичку надградњу која побољшава проналажење информација са различитих локација на мрежи, повезује их са оквирима који се базирају на технологијама семантичког веба. Семантички веб као мрежа података захтева да се подаци постављају у одређеним стандардним форматима да би технологије уграђене у семантички веб могле да се примене.

Посебан део семантичког веба који користе модели метаподатака у Еуропеани јесте Linked (Open) Data тј. технологији повезаних података.²⁴⁹ Linked Open Data или Повезани отворени подаци је заправо база међусобно повезаних скупова стандардизованих података. Ти подаци су структурирани на посебан начин тако да омогућавају да се метаподаци повезују и допуњавају. Тиме се омогућава да се пронађу различита представљања истог садржаја, као и да се направе везе међу блиским садржајима односно ресурсима.

Пројекат Linked Open Data успостављен је 2007. године, а подржала га је W3C Semantic Web Education and Outreach Interest Group.²⁵⁰ У првој фази установљени су скупови података који су постављени за слободно коришћење (отворени приступ). Ти подаци су конвертовани у RDF према правилима повезаних података и постављени су на мрежу. У тој раној фази у пројекту су учествовали истраживачи и програмери из универзитетских истраживачких лабораторија. Укључивањем у пројекат великих

²⁴⁷OWL – Web Ontology Language, доступно на: <http://www.w3.org/2001/sw/wiki/OWL>

²⁴⁸ „У контексту рачунарских и информационих наука онтологија дефинише скуп репрезентативних појмова са којима се врши моделовање у домену знања или дискурса. Репрезентативни појмови су обично класе (или скупови), атрибути (или својства) и релације (или односи између чланова класе).” – Tom Gruber. Ontology. In: Ling Liu and M. Tamer Özsu (eds.) *Encyclopedia of Database Systems* (Heidelberg etc.: Springer-Verlag, 2009) <http://tomgruber.org/writing/ontology-definition-2007.htm> (приступљено 22.10.2015) (превод аутора)

²⁴⁹Linked (Open) Data, доступно на: <http://linkeddata.org/home>

²⁵⁰W3C Semantic Web Education and Outreach (SWEO) Interest Group, доступно на: <http://www.w3.org/blog/SWEO/>

институција као што су BBC²⁵¹, Thomson Reuters²⁵² и Конгресна библиотека²⁵³, број повезаних података се значајно повећао, а то је и тренутак када су библиотеке спознале могућност и корист овакве иницијативе. За Linked Open Data изузетно је значајно учешће библиотека са својим подацима јер су подаци у библиотекама углавном у стандардним форматима који се једноставно конвертују, а подаци се онда лако повезују и преносе.

Битно је истаћи да повезани подаци могу садржати било који тип података и да Linked Open Data омогућава да свако може да постави и повеже своје податке са већ постојећим подацима. Linked Open Data обједињује много различитих система и метаподатака и представља мрежу међусобно повезаних скупова података. Свеукупност повезаних скупова података представљена је на слици 20. За повезивање ентитета користе се RDF линкови помоћу којих се ствара општи дијаграм који обухвата различите изворе података и омогућава откривање нових извора. У центру дијаграма је DBpedia²⁵⁴ која екстрахује структуриране информације са Википедије и обједињује и повезује те податке са другим изворима.

Уколико погледамо поменути дијаграм, приметимо да су на њему скупови података представљени различитим бојама: сивом бојом извори општих података, тиркизном бојом извори друштвених података, жутом бојом извори географских и статистичких података, љубичастом бојом извори мултимедијалних података, розе бојом извори података везаних за медицину и здравље, светло зеленом бојом извори библиографских података итд. Илустрације ради, поменућемо најзначајније библиографске изворе – Open Library²⁵⁵, LCSH²⁵⁶, RAMEAU²⁵⁷, MARC Code Lists²⁵⁸, data.bnf.fr²⁵⁹, GND²⁶⁰, VIAF²⁶¹, Deutsche Biographie²⁶², GeoNames²⁶³ и сл.

²⁵¹BBC – British Broadcasting Corporation, доступно на: <http://www.bbc.com/>

²⁵²Thomson Reuters, доступно на: <http://thomsonreuters.com/en.html>

²⁵³Library of Congress, доступно на: <http://www.loc.gov/>

²⁵⁴DBpedia, доступно на: <http://dbpedia.org/About>

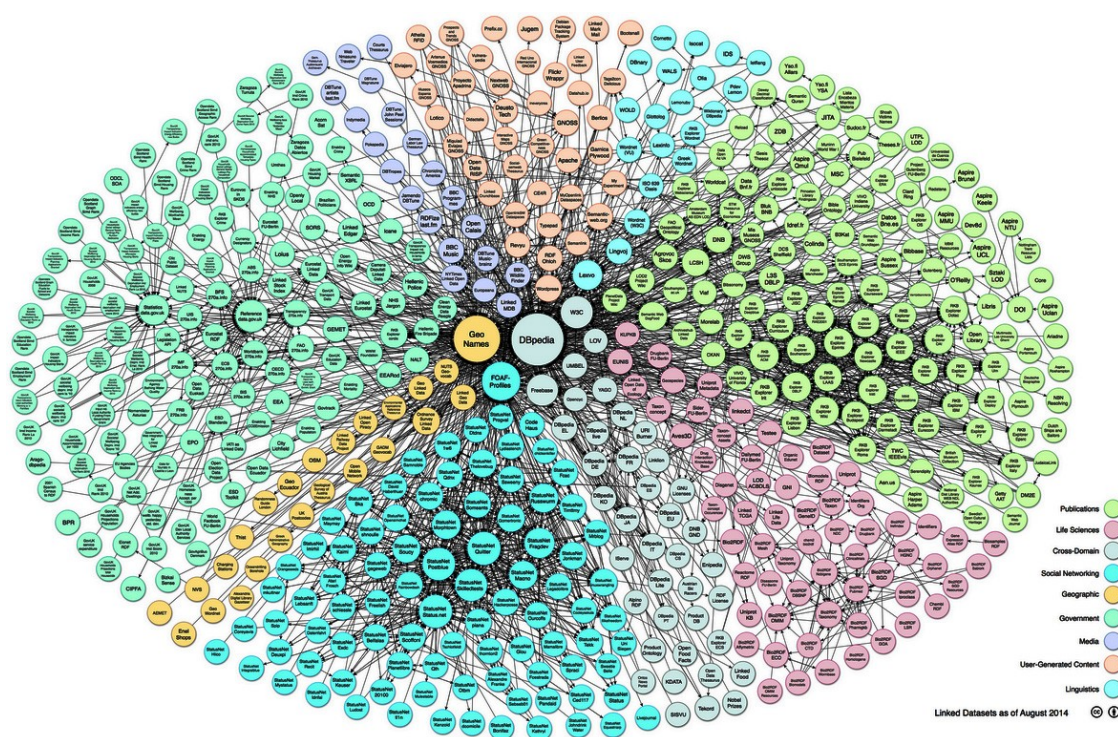
²⁵⁵Open Library, доступно на: <https://openlibrary.org/>

²⁵⁶ LCSH – Library of Congress Subject Headings, доступно на: <http://id.loc.gov/authorities/subjects.html>

²⁵⁷RAMEAU Répertoire d'autorité-matière encyclopédique et alphabétique unifié, доступно на: <http://rameau.bnf.fr/>

²⁵⁸Marc Code Lists, доступно на: <http://www.loc.gov/marc/>

²⁵⁹Data about BnF, доступно на: <http://data.bnf.fr/>



Слика 20 – Linked Open Data – дијаграм међусобно повезаних скупова података²⁶⁴

Основне технологије на којима се базира технологија повезаних података су URI као опште средство за идентификацију ентитета или концепата, односно јединствени идентификатор²⁶⁵, HTTP као једноставан, али универзалан начин за проналажење и приказ ресурса и RDF као модел података којим се структурирају и повезују описани подаци.²⁶⁶

Бернерс-Ли²⁶⁷ сматра да технологија повезаних података на семантичком вебу треба да поштује неколико принципа (заснованих на повезивању):

²⁶⁰Gemeinsame Normdatei (The Integrated Authority File), доступно на: http://www.dnb.de/EN/Standardisierung/GND/gnd_node.html

²⁶¹Virtual International Authority File, доступно на: www.viaf.org

²⁶² Deutsche Biographie, доступно на: <http://www.deutsche-biographie.de/>

²⁶³GeoNames, доступно на: <http://www.geonames.org/>

²⁶⁴Richard Cyganiak and Anja Jentzsch. „Linking Open Data cloud diagram“, <http://lod-cloud.net> (приступљено 10.09.2014)

²⁶⁵У смислу традиционалне библиотеке близак функцији идентификатора у нормативним датотекама

²⁶⁶Christian Bizer, Tom Heath, Tim Berners-Lee. „Linked Data – the Story So Far.“ *Linked Data, International Journal on Semantic Web and Information Systems*, special issue (2009), <http://tomheath.com/papers/bizer-heath-berners-lee-ijswis-linked-data.pdf> (приступљено 23.10.2015)

²⁶⁷Tim Berners-Lee. „Linked Data.“ World Wide Web Consortium, 2006, <http://www.w3.org/DesignIssues/LinkedData.html> (приступљено 10.09.2014)

- Користити јединствене идентификаторе извора (URI) за именовање појмова
- Користити HTTP URI тако да људи буду у стању да претражују имена појмова
- Када неко претражује одређени URI, обезбедити му и друге корисне информације, коришћењем дефинисаних стандарда (нпр. RDF)
- Укључити линкове ка другим URI да би се могло открити више.

Еуропеана у свим својим пројектима инсистира и на принципима повезаних података онако како их је дефинисао Бернерс-Ли, али и на стандардним форматима метаподатака и дигиталних објеката. Као што је раније поменуто, ти стандардни формати у којима установе достављају метаподатке могу бити Даблинско језгро или MARC 21 формат који се, према задатим циљевима конкретног пројекта, касније трансформишу у проширене моделе метаподатака ESE или EDM.

Еуропеана записе на сематичком вебу у првом реду базира на везама са следећим скуповима података:

- VIAF– међународној виртуелној нормативној датотеци²⁶⁸²⁶⁹ (имена аутора која се јављају у записима у Еуропеани у етикети **author** се путем јединственог идентификатора извора – URI повезују са записима у VIAF; на тај начин постиже се да се претрагом неког аутора у оквиру Еуропеане и везом ка VIAF-у дође и до осталих података о аутору и његовим делима у другим системима који своје податке повезују са базом VIAF);
- GND– интегрисаној нормативној датотеци имена аутора, предметних одредница и корпоративних тела коју развија Немачка национална библиотека (имена аутора која се не нађу у бази VIAF, потраже се у овој бази и подаци се повежу по истом принципу и са истим циљем; као и предметне одреднице и називи корпоративних тела);

²⁶⁸ „Нормативна датотека је резултат нормативне контроле која доприноси Универзалној библиографској контроли. То је библиографска база јединствено унетих података, пре свега личних имена аутора, чиме се омогућава да у узајамном каталогу, на пример, сва дела једног аутора буду окупљена око унифицирано унете одреднице, чиме се пружа прилика за мерење фреквенције коришћења, али и чувања ауторских права.“ – Александра Вранеш и Љиљана Марковић. *Од рукописа до библиотеке*. Нормативна датотека (Београд: Филолошки факултет, 2008), 235.

²⁶⁹ „Нормативна датотека је списак усвојених облика одредница које се користе у каталогу библиотеке или датотеке библиографских записа, а одржавање датотетке омогућава да се одредница доследно примењује и да се нови облици додају колекцији. Посебне нормативне датотеке се воде за имена аутора, јединствене наслове, наслове колекција и предметне одреднице.“ – Joan M. Reitz. ODLIS – Online Dictionary for Library and Information Science, Authority file, <http://www.abc-clio.com/ODLIS/searchODLIS.aspx> (приступљено 22.10.2015) (превод аутора)

- LCSH– нормативној датотеци предметних одредница Конгресне библиотеке (метаподаци који се налазе у записима Европеане у етикетама **subject**, путем јединственог идентификатора извора – URI повезују се са записима у LCSH што доводи до бољег проналажења информација у дигиталним објектима чији садржај се бави истим предметом);
- GeoNames– географској бази података која садржи преко 10 милиона географских имена (метаподаци о географским појмовима који се у записима Европеане налазе у етикетама **coverage** повезују се са подацима у бази GeoNames који, осим географских имена места на различитим језицима, садрже најразличитије географске податке укључујући податке о површинама, надморским висинама, дубинама, становништву и сл.).

Сви описани системи (COBIB.SR, PHAIDRA, Дигитална библиотека Филолошког факултета, Европеана) поседују своје схеме и системе метаподатака за опис дигиталних објеката. Ако упоредимо линеране моделе података као што су јавно доступни каталози библиотека (COBIB.SR) са разгранатом семантичком структуром метаподатака (попут ове у записима Европеане) можемо претпоставити да ће структурирани метаподаци повезани са мрежом семантичких података допринети бољем проналажењу информација. Анализа примера проналажења информација биће дата у поглављу 5 где ћемо показати колико је прецизан и квалитетан одзив на постављене упите, на који начин се пронађене информације везују са другим сличним изворима, као и да је почетна претпоставка тачна.

4.3 Дескриптори садржаја

4.3.1 Ручно и аутоматско индексирање – разлике

Да би садржина неког документа била претражива и да би претраживање било успешно, неопходно је да тај документ буде индексиран. Индексирање као резултат анализе садржаја документа (и штампаног и дигиталног) јесте поступак путем кога се неком извору информација придружују индексни термини, односно идентификатори,²⁷⁰ односно индексирање је процес конструисања сурогата документа придруживањем идентификатора текстуалним јединицама.²⁷¹ Индексирање може обављати човек, стручњак из одређене области – у том случају говоримо о ручном

²⁷⁰John Feather and Paul Sturges (eds.) *International Encyclopedia of Library and Information Science. Indexing.* (London: Routledge, 2003), 236.

²⁷¹Cvetana Krstev. *Materijali za kurs Pronalaženje informacija, deo Automatsko indeksiranje*, nepublikovani materijal za školsku 2015/2016.

индексирању или машина, аутоматски подржана процедура – у том случају реч је о аутоматском индексирању.

Оба случаја подразумевају доделу идентификатора садржаја која корисницима омогућава да преко неког конкретног упита добију одговор у виду одговарајућег документа. Ручно индексирање није мање квалитетно, али оно може обухватити мање колекције докумената у задатом временском интервалу у односу на аутоматско индексирање.

Као резултат обе врсте индексирања добијамо индексне речнике: контролисане (углавном добијене ручним индексирањем коришћењем термилошких листа и поштовањем правила за речничке форме и коришћење термина) и неконтролисане (углавном у окружењу аутоматског индексирања). Процедуре аутоматског индексирања углавном су засноване на руковању текстом информационих јединица, тако да разликујемо: индексирање додељивањем термина и индексирање екстраховањем термина из текста. У другом случају индексни речник је много слабије контролисан него код ручног индексирања које се махом своди на индексирање додељивањем термина. Индексирање се може вршити једним термином или фразним терминима. Када користимо индексирање једним термином подразумева се да се „индексне јединице које се каче уз информационе јединице састоје од скупа појединачних термина од којих сваки одражава неки аспект садржаја текста, док индикатори веза међу тим терминима у индексној јединици не постоје.“²⁷²

Методe индексирања могу бити пре-координисано и пост-координисано индексирање. Код пре-координисаног индексирања документима се придружују комплексне фразе или групе термина са спецификованим назнакама веза које међу њима постоје, док се код пост-координисаног индексирања терминима придружују референсни термини као што су „види“ и „види и.“

Осим што ручним индексирањем није могуће анализирати велике колекције докумената, оно је и субјективно јер тешко да ће два стручњака, индексера, доделити идентичне индексне термине једном документу. Ако имамо у виду само око 400.000 страница српских историјских новина у пројекту „Европске новине“ јасно је да ручним индексирањем не бисмо далеко стигли, у најбољем случају индексирали

²⁷² Cvetana Krstev. *Materijali za kurs Pronalaženje informacija, deo Automatsko indeksiranje*, nepublikovani materijal za školsku 2014/2015.

бисмо један наслов ангажовањем више стручњака индексера, али тиме бисмо ризиковали неуједначеност у погледу доделе индексних термина.

Аутоматско индексирање је ослобођено субјективности јер се поступак врши на идентичан начин за сваки документ, а укључује фреквенцију појављивања термина (речи)²⁷³ у одређеном тексту при чему се поштују принципи идентификовања индивидуалних речи у тексту – лексичка анализа, уклањање функционалних речи и високофреквентних термина јер су недовољно специфични у погледу садржинске анализе применом искључне листе²⁷⁴, свођење преосталих речи на њихов корен или речнички облик – лему, опционо формирање фраза и индексних термина, опциона замена речи, корена речи или фразе њиховом класом термина из тезауруса, прорачунавање вредности сваког преосталог корена речи или речи, класе термина из тезауруса или фразног термина.²⁷⁵

Салтон сматра да се предности аутоматског индексирања огледају у томе што је могуће постићи задовољавајући ниво конзистентности индексирања, цена производње индексних јединица је мања ако се у обзир узме дужи временски период, смањује се време потребно за индексирање и може се постићи боља ефикасност проналажења.²⁷⁶ Овим предностима аутоматског индексирања можемо додати и факторе објективности и могућност индексирања великих колекција докумената.

Сумирање разлика између ручног и аутоматског индексирања свешћемо на то да се код ручног индексирања чешће користе контролисани индексни речници, додељивање индексних јединица, фразни термини и пре-координисано индексирање са терминима у контексту, док се код аутоматског индексирања користе чешће неконтролисани индексни речници, екстраховање термина из текста документа, појединачни термини и пост-координисано индексирање.

4.3.2 Одзив, прецизност и фреквенција појављивања термина

За проналажење информација и аутоматско индексирање изузетно су значајни појмови „одзив”, „прецизност” и „фреквенција термина.” Будући да се у овом

²⁷³ Дефиниција појма и додатна објашњења биће дата у наредном поглављу

²⁷⁴ Искључна листа је листа речи које су недовољно специфичне и изостављају се из садржинске анализе текста или неког аутоматског процесирања података (обично су у питању чланови, везници, речце и сл.)

²⁷⁵ Marie-Francine Moens. *Automatic Indexing and Abstracting of Document Texts*. (Dordrecht: Kluwer Academic Publishers, 2000), 94

²⁷⁶ Gerard Salton. *Automatic Text Processing – The Transformation, Analysis, and Retrieval of Information By Computer*. (Boston: Addison-Wesley Publishing Company, 1989)

истраживању у домену проналажења информација базирамо на ефикасности претраживања²⁷⁷ где се ефикасност проналажења информација мери односом одзива и прецизности, нужно је ближе објаснити поменуте појмове и њихов однос.

Одзив (R од енг. recall) представља однос између пронађеног релевантног материјала и укупно релевантног материјала, а дефинише се као:

$$R = \frac{\text{број пронађених релевантних јединица}}{\text{укупан број релевантних јединица у колекцији.}}$$

Прецизност (P од енг. precision) је однос пронађеног релевантног материјала према укупно пронађеном материјалу, а дефинише се као:

$$P = \frac{\text{број пронађених релевантних јединица}}{\text{укупан број пронађених јединица.}}^{278}$$

Према Салтону, одзив и прецизност крећу се од 0 до 1 што је еквивалентно вредностима од 0 до 100%. Сваки корисник жели да претраживањем постигне и велики одзив и велику прецизност, односно жели да пронађе велики део корисног материјала, а да истовремено велики део пронађених јединица буде релевантан. Међутим, у пракси је неопходно успостављање компромиса јер се истовремена оптимизација одзива и прецизности не може постићи. Тако, ако корисник бира између велике специфичности и велике ширине термина, боље је да се определи за специфичност. Велика ширина термина доводи до високог одзива, а ниске прецизности, па корисник може добити велику количину пронађеног материјала међу којим је и много нерелевантног. Висока прецизност резултира мањим бројем јединица које је лакше испитати. Лакше је поправити недостатак прецизности јер се прецизност може измери тиме колика је корисност пронађеног материјала, а одзив зависи и од релевантних јединица колекције које нису пронађене.

Фреквенција термина у датом документу јесте број који означава колико пута се тај термин појављује у том документу. Аутоматско пребројавање одређеног термина у неком тексту као методу анализе предмета тог документа успоставио је Ханс Петер Лун (Hans Peter Luhn) који је сматрао да:

²⁷⁷Marie-Francine Moens. *Automatic Indexing and Abstracting of Document Texts*. (Boston: Kluwer Academic Publishers, 2000), 78

²⁷⁸Gerard Salton. *Automatic Text Processing – The Transformation, Analysis, and Retrieval of Information By Computer*. (Boston: Addison-Wesley Publishing Company, 1989), 277-278

- „фреквенција појављивања речи у неком документу даје корисну меру значаја те речи;
- релативна позиција речи унутар реченице даје корисну меру за одређивање значаја те реченице; и
- фактор значаја неке реченице засниваће се на комбинацији ова два захтева.”²⁷⁹

Када говоримо о фреквенцији појављивања термина (речи) треба истаћи да граматичке функционалне речи као што су у енглеском *and*, *or*, *of* или *but* имају приближно исту фреквенцију појављивања у свим документима колекције и то углавном високу фреквенцију. Одређивање оних термина који су битни за садржај документа и дефинисање њихове важности (мерење тежине термина) је кључно у процесу индексирања и квалитетном проналажењу релевантних докумената. Тежина термина је, према Луну, сразмерна фреквенцији термина. Нефункционалне речи (термини) имају много нижу фреквенцију појављивања у различитим текстовима колекције. Управо се нефункционалне речи могу односити на садржај документа и њихова фреквенција појављивања може да указује на значај тог термина за представљање садржаја документа.

Ипак, мерење фреквентности неког термина задовољава само један од циљева проналажења информација – одзив. Уколико тражимо неки појам и добијемо високу фреквенцију појављивања тог појма у документима, добили смо велики одзив на постављени упит. То ипак не значи да је и прецизност висока јер висока фреквенција појављивања термина у неком документу не значи да је тај документ релевантан за наше претраживање (битан је контекст појављивања термина, корелација са другим терминима и сл.). Такође, прецизности може сметати и то што ће високим одзивом бити пронађено доста истих, а не различитих докумената. Другу врсту проблема представља ситуација кад се термин појављује у многим документима, па не можемо да их разликујемо један од другог. Функција прецизности много боље се задовољава ако се термин појављује ретко, само у појединим документима, јер је на тај начин лако издвојити документа у којима се термин појављује од оних докумената у којима се не појављује.

²⁷⁹Hans Peter Luhn. „The automatic creation of literature abstracts.” *IBM Journal of Research and Development* 2(1958):159-165.

Да би аутоматско индексирање било квалитетно, да бисмо извршили поуздано проналажење информација са високим одзивом и прецизношћу, као и да би се анализа лексичким ресурсима и алатима исцрпно спровела, неопходно је квалитетно аутоматски обрадити документе који ће се анализирати. То се постиже применом технологија за пречишћавање текста.

4.3.3 Технологије за пречишћавање текста

Како је раније објашњено, метаподацима се описују различите особине дигиталних објеката. Осим формалних карактеристика, метаподацима се донекле открива и садржај дигиталног објекта јер се у етикетама subject, coverage, abstract и сл. наводе и кључне речи или предметне одреднице и ознаке, као и сажетци које указују на садржај. Као што смо видели на анализираним примерима, некада су те кључне речи и предметне одреднице прецизно дефинисане, некада мање прецизно, па одзив у проналажењу информација зависи и од вештине особе која описује дигитални објекат.

Ради повећања прецизности и скраћивања времена потребног да се пронађе релевантан дигитални објекат или релевантна информација садржана у њему, осим метаподатака, у дисертацији ћемо анализирати и различите технологије за пречишћавање текста и утврдити у коликој мери оне доприносе квалитетнијем и тачнијем проналажењу информација у дигиталним библиотекама. Технологије за препознавање информација садржаних у објектима дигиталних библиотека су:

a. Оптичко препознавање карактера – OCR (Optical Character Recognition)

Дигиталне библиотеке се састоје од дигиталних објеката које је могуће претраживати и путем метаподатака и преко целог садржаја текста, али и идентификовати слике, фотографије и разне мултимедијалне садржаје. Једноставним скенирањем неког документа добијамо његову „слику”, чак и кад је реч о изворно текстуалним садржајима. Погрешно је становиште да је дигитализација једног документа завршена његовим скенирањем и превођењем у „слику.” Ту је заправо почетак дигитализације јер такве „слике” докумената можемо листати, прегледати, читати, али тиме дигитална библиотека није у великој предности у односу на традиционалну библиотеку, осим што је приступ садржајима могућ са удаљеног места.

Квалитетна дигитализација штампаног или рукописног текста или сликовне грађе подразумева следеће фазе које су део претходног процесирања материјала након кога следи оптичко препознавање карактера :

- Скенирање штампаног материјала (превођење у стандардне формате – препоручује се TIFF формат јер JPEG и PNG често као излаз дају мутне и недовољно јасне слике);
- Обрада слике (раздвајање страница, исецање делова и поравнавање маргина; уклањање закривљења насталих процесом скенирања, прилагођавање контраста, прилагођавање величине и резолуције слике);
- Креирање излазног документа који ће се користити у дигиталним библиотекама (формат у коме тај документ треба да се појави је најчешће PDF на коме се касније примењују технологије оптичког препознавања карактера).²⁸⁰

Да бисмо поспешили претраживања података и лакше дошли до жељеног дигиталног садржаја, користи се поменута метода оптичког препознавања карактера. Оптичко препознавање карактера је конверзија, односно превођење руком писаног текста, текста куцаног на машини или штампаног текста у машински кодиран текст. Од првих система који су вршили препознавање карактера слово по слово где је текст представљен као ниска карактера, данас су развијени системи који препознају и руком писан курзив, па чак и код оних језика код којих нема размака између речи. Дигиталне библиотеке садрже дигиталне објекте на које су примењене ове технологије јер једино тако могу понудити претраживање преко сваке појединачне речи у тексту (осим у случајевима кад је текст прекуцаван или је изворно дигиталан).

Укратко ћемо описати основне кораке у процесу оптичког препознавања карактера:

- Учитавање слике извора који се обрађује – добар систем за оптичко преознавање карактера мора да подржи различите формате за дигиталне фотографије скенираних материјала – (поменути формати – TIFF, JPEG, PNG.) Датотеке у PDF формату такође морају бити подржане јер се преко њих најчешће и врши оптичко препознавање карактера;

²⁸⁰ Nikola Ikonov and Milena Dobrea. „The making of... digital book.” *NCD Review* 13(2008):1-8, , <http://elib.mi.sanu.ac.rs/files/journals/ncd/13/ncd13001.pdf> (приступљено 11.09.2014)

- Утврђивање резолуције слике и врсте и величине фонтова који се јављају у тексту;
- „Бинаризација” или претварање колор фотографија скенираног извора у црно-беле фотографије са добрим контрастом;
- Уклањање оквира и подвлака у тексту;
- Утврђивање и означавање речи и размака у тексту;
- Препознавање знакова као главног дела оптичког препознавања карактера. Слика сваког карактера конвертује се у одговарајући код карактера према учитаним кодним табелама. Препознавање никада није 100% успешно, па се код непрепознатих карактера поправке најчешће врше ручно на основу извора. Најбољи су они системи који таквим на поправкама „уче“ тј. програмирани су тако да се „обучавају” на одређеном броју страница у току чега корисник даје разрешења за све недоумице или систем усваја промену и извршава је у целокупном тексту након ручне поправке на погрешно препознатом карактеру;
- Коришћење речника као подршке за боље препознавање карактера. Овај корак може да побољша квалитет препознавања, неки карактери могу да изгледају слично, али појављивање одређене речи у електронском речнику помаже код тачнијег оптичког препознавања;
- Чување резултата у различитим излазним форматима – софтвер мора понудити следеће формате – PDF, .doc, .txt. Нарочито је пожељан PDF формат јер је најшире распрострањен у дигиталним библиотекама.

Овако припремљени дигитални објекти омогућавају претрагу целокупног текста документа. То је уједно огромна предност дигиталне библиотеке у односу на традиционалну и у погледу приступа садржајима, и у погледу проналажења информација. У истраживању технологија за оптичко препознавање карактера које су широко доступне на вебу, запажа се да су најбројнији они алати за оптичко препознавање карактера штампаних латиничних фонтова, па и рукописних текстова писаних латиницом, као и ћириличних писама и фонтова.

У ресурсима који се анализирају у овој докторској дисертацији, технологија оптичког препознавања карактера се примењује на дигиталним објектима Дигиталног репозиторијума Универзитета у Београду, Дигиталне библиотеке Филолошког факултета и дигиталним материјалима у оквиру Еуропеана пројеката.

6. Оптичка сегментација чланака и класе странице – OLR (Optical Layout Recognition)

Још једна технологија је изузетно значајна за квалитетније препознавање карактеристика и садржаја дигиталних објеката, па самим тим и за проналажење информација о њима. Реч је о оптичкој сегментацији чланака. Важност ове технологије је нарочито видљива код аутоматске обраде новинских колекција и колекција часописа, јер она на неки начин представља пут ка аутоматизованој каталогизацији ових извора, а подразумева да је дигитални објекат најпре обрађен технологијом оптичког препознавања карактера. Односи се на аутоматску доделу структуралних и библиографских метаподатака о чланцима, пре свега оних из новинских колекција.

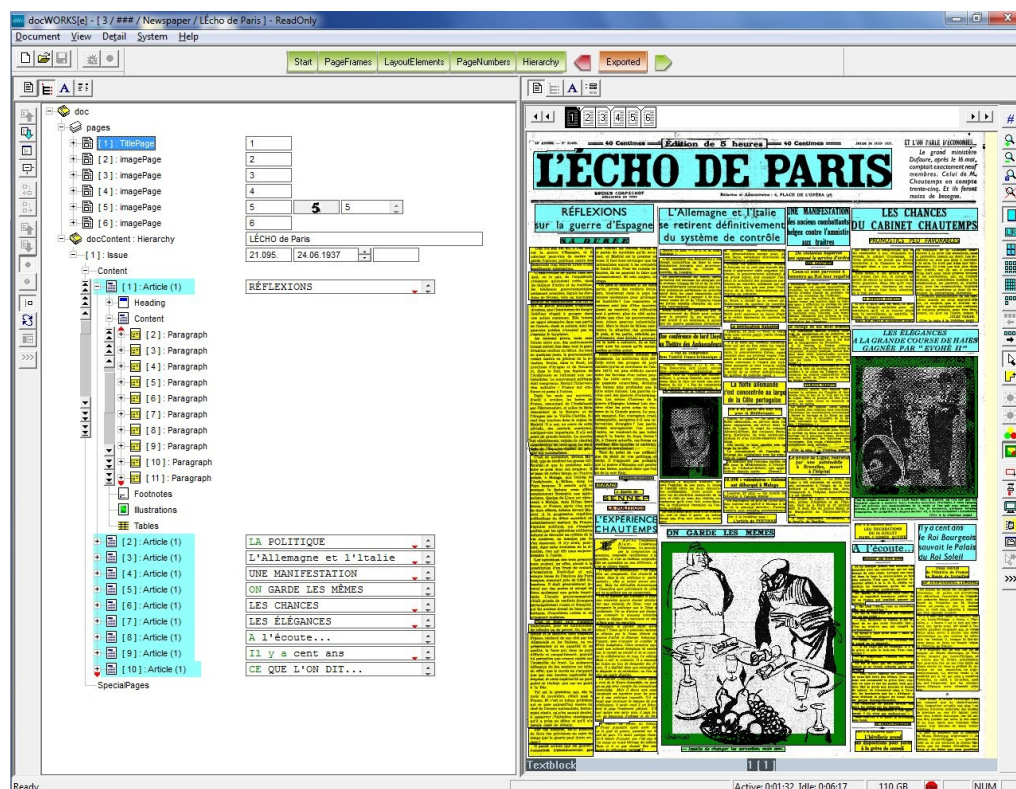
Оно што је битно код аутоматске доделе ових метаподатака јесте то да је неопходан предуслов квалитетна дигитализација новинских страница и страница часописа у свим корацима поменутих у претходном поглављу. Једино тако ће софтвер који препознаје податке са високим процентом тачности моћи да детектује назив серијске публикације, наслов чланка, број, датум, годину, годиште и бројеве страница на којима се чланак налази.

Помоћу технологије оптичке сегментације чланака и класе странице могуће је препознавати речи, колоне са текстом, текстуалне блокове, рубрике, илустрације, рекламе и огласе, табеле, као и следеће типове страница: насловну страну, садржај, странице са илустрацијама (оне које садрже бар једну илустрацију), као и странице које садрже искључиво рекламе и огласе. Такође, могуће је препознавање наставака чланака (уколико чланак почиње на једној, а завршава се на другој страни, као и прелазак чланка из једне у другу колону). За сваки од чланака препознатих путем оптичке сегментације, аутоматски се врши додела метаподатака у жељеном, претходно задатом формату.

Када су оптички препознати наслови чланака, потписи аутора, наслов серијске публикације, број, годиште, број стране и када је могуће те податке, односно метаподатке сместити у одређени стандардизовани формат који комуницира са другим форматима и системима, онда је квалитет дигиталне библиотеке повећан, смањен је посао на тзв. „ручном” уносу, а могуће је обрадити велике колекције докумената. Једино на овај начин је могуће урадити комплетну аналитичку обраду новинских колекција.

За разлику од традиционалне библиотеке, где су новинске колекције и колекције часописа доступне у штампаном или микрофилмованом облику и могу се прегледати и прелиставати, дигитална библиотека нуди претрагу преко сваке речи у чланку и метаподатака о њему, а коришћењем технологије за оптичку сегментацију чланака и класе странице поступак доделе метаподатака је у великој мери поједностављен.

У оквиру ресурса које у овом раду користимо за анализу, технологија оптичке сегментације чланака и класе странице користи се у пројекту „Европске новине.” На слици 21 дат је приказ алата који врши оптичку сегментацију чланака. Плавом бојом аутоматски су пребојени наслов новина, наслови чланака и њихови аутори и из њих добијени библиографски метаподаци. Текст чланака је пребојен жутом бојом, а слике и објашњења уз слике зеленом бојом.



Слика 21 – Приказ оптичке сегментације чланака и аутоматског издвајања метаподатака (новине L'echo de Paris, за 24. јун 1937.)²⁸¹

Узимајући у обзир све наведене карактеристике оптичке сегментације чланака, сматрамо да ова технологија позитивно утиче на проналажење информација у дигиталним библиотекама, али у контексту истраживаних колекција није примењена.

²⁸¹Слика преузета са веб странице Европеане и пројекта Европске новине, <http://www.europeana-newspapers.eu/activities/> (приступљено 28.10.2015)

в. Препознавање именованих ентитета – NER (Named Entity Recognition)

Једна од технологија које се користе за проналажење информација у самим документима које се обавља коришћењем метода екстракције информација је препознавање именованих ентитета. Ова технологија је намењена лоцирању појмова у тексту и њиховом класификовању у одређене категорије. Први скуп именованих ентитета садржао је следеће категорије: лична имена, називи организација, географски појмови (локације), временске категорије, новчани износи, одређени типови нумеричких израза (нпр. мерни изрази).²⁸² Првобитно утврђене категорије именованих ентитета разлагане су на саставне делове (нпр. лично име на презиме, женско име, мушко име), а додаване су и нове категорије како се појављивала потреба за њима у конкретним истраживањима (нпр. називи производа, објеката, историјских догађаја, награда, геополитичких територија и сл.). У раду Секина и др. из 2002. дат је преглед за преко 150 категорија именованих ентитета са примерима (нпр. име – организација – компанија – пример IBM, Microsoft),²⁸³ а у раду из 2004. преглед за преко 200 категорија у хијерархијској структури дрвета (нпр. име – локација – геополитичка јединица – град).²⁸⁴

Код малих колекција могуће је ручно означити основне категорије именованих ентитета. Међутим, кад постоји овако велики број категорија наведен у претходном пасусу и код великих колекција докумената, какве су дигиталне библиотеке, тај процес се обавља аутоматски. За креирање ових аутоматизованих система за српски језик користе се модели који су засновани на лингвистичким знањима и на особинама конкретног природног језика. Морфолошки систем српског језика, као и морфолошки системи других словенских језика, изузетно је богат, па тако велики број различитих

²⁸²Ralph Grishman and Beth Sundheim. „Message Understanding Conference 6: A Brief History”. In: COLING-1996, Copenhagen, Denmark, pp. 466-471, , <http://www.aclweb.org/anthology/C96-1079> (приступљено 04.11.2015)

²⁸³Satoshi Sekine, Kiyoshi Sudo and Chikashi Nobata. „Extended Named Entity Hierarchy.” In: *Proceedings of the Third International Conference on Language Resources and Evaluation (LREC '02)*, Las Palmas, Canary Islands, Spain, May 29-31 2002, pp. 1818-1824, <http://www.lrec-conf.org/proceedings/lrec2002/pdf/120.pdf> (приступљено 04.11.2015)

²⁸⁴Satoshi Sekine and Chikashi Nobata. „Definition, dictionaries and tagger for Extended Named Entity Hierarchy.” In: *Proceedings of the Forth International Conference on Language Resources and Evaluation (LREC '04)*, Lisbon, Portugal, May 26-28 2004, pp. 1977-1980, <http://www.lrec-conf.org/proceedings/lrec2004/> (приступљено 04.11.2015)

морфолошких облика учествује у систему за препознавање именованих ентитета.²⁸⁵ Технологију за препознавање именованих ентитета у српском језику развија Група за језичке технологије Математичког факултета Универзитета у Београду. У раду учествују и сарадници са других факултета овог Универзитета, у првом реду Филолошког.

Препознавање именованих ентитета у дигиталним библиотекама различитих докумената на нашем језику у тесној је вези са развојем електронских морфолошких речника српског језика, базираним на свим његовим специфичним карактеристикама. Нарочито су значајни они сегменти морфолошких речника српског језика који су у вези са личним именима, геополитичким именима, именима организација и сл. Екстракција информација из дигиталног документа коришћењем технологије препознавања именованих ентитета треба да омогући повезивање препознатог ентитета са другим ресурсима на мрежи. Раније су поменуте различите нормативне датотетке личних, географских имена и сл. које су део ширег семантичког веба Повезани (отворени) подаци, а са којима се може извршити повезивање термина након препознавања именованих ентитета (нарочито за властита имена, имена организација и географских појмова).

Препознавање именованих ентитета у употреби је у Европеана пројектима, нарочито у пројекту „Европске новине.” Нажалост, у оквиру овог пројекта није предвиђено препознавање именованих ентитета за српски језик, већ само за енглески, немачки и холандски.

Циљ примене технологије препознавања именованих ентитета на предвиђеном корпусу текстова и објеката дигиталних библиотека је да се испитају могућности екстракције информација из дигиталних докумената применом наведене технологије и на пронађеним примерима анализирају предности оваквог приступа у односу на традиционалне методе.

Како је препознавање именованих ентитета у српском језику (NER за српски језик) уско повезано са електронским речницима простих и сложених речи,²⁸⁶ у наредним одељцима описаћемо лексичке ресурсе и алате које развија Група за језичке технологије.

²⁸⁵Cvetana Krstev, Ivan Obradović, Miloš Utvić and Duško Vitas. „A System for Named Entity Recognition Based on Local Grammars.” *Journal of Logic and Computation* 24, 2(2014):473

²⁸⁶Cvetana Krstev, Ivan Obradović, Miloš Utvić and Duško Vitas. „A System for Named Entity Recognition Based on Local Grammars.” *Journal of Logic and Computation* 24, 2(2014):477

4.4 Лексички ресурси и алати

Да би се методе проналажења и екстракције информација у дигиталним библиотекама квалитетно спровеле, често се користе одређени лексички ресурси и алати. Нарочито треба имати у виду вишејезичност дигиталних докумената и потребу да се постављање упита врши на што једноставнији начин. Како се у овој дисертацији анализирају, у првом реду, дигиталне библиотеке које садрже дигиталне објекте на српском и енглеском језику, то су и лексички ресурси и алати који се описују и користе развијани управо за ове језике.

Корпус савременог српског језика (СрпКор)

Електронски корпус једног језика може се дефинисати као „колекција аутентичних машински читљивих текстова који представљају репрезентативни узорак једног језика или језичког варијетета.”²⁸⁷ Корпус савременог српског језика (СрпКор) развија Група за језичке технологије на Математичком факултету Универзитета у Београду, а његова прва верзија Неетикетирани корпус српског језика (НЕТК) доступан је онлајн од 2003. године. У том тренутку то је била неанотирана колекција текстова са 22,2 милиона речи без информација о изворима. У тој првој фази радило се на прикупљању текстова за корпус, тестирању алата за креирање корпуса и његову претрагу, изради базе података и веб странице преко које би корисници приступали корпусу, одакле би се вршила администрација корпуса и омогућила једноставна претрага.

Током постојања овај корпус је постепено мењао свој изглед и функционалност. Најпре је допуњен информацијама о изворима, а корисницима је омогућен увид у библиографске референце текстова из којих су издвајане конкорданце.²⁸⁸ У периоду од 2005. до 2010. године Група за језичке технологије Математичког факултета која ради на развоју СрпКор-а учествовала је у пројектима који су омогућили конструкцију паралелизованих корпуса и развој потребних алата, проширивање електронског морфолошког речника сложеним речима и развој лексичких ресурса за препознавање именованих ентитета. Од 2011. до 2013. СрпКор је ажуриран три пута, да би 2013. корпус достигао 122 милиона корпусних речи са

²⁸⁷ Anthony McEnery, Richard Xiao and Yukio Tono. *Corpus Based Language Studies: an Advance Resource Book*. (London: Routledge, 2006),15.

²⁸⁸ Конкорданца је алфабетски уређена листа речи у тексту или групи текстова заједно са позицијама у тексту сваког појављивања речи, као и цитатима или изводима из контекста – према Keith Brown (ed.) *The Encyclopedia of Language and Linguistics*. Concordance. (Oxford: Elsevier, 2005), 182.

аутоматским морфолошким анотацијама. СрпКор данас броји око 5.000 текстова који су настали од 1910. до 2012. године при чему су најбројнији текстови изворно настали од 2000. до 2009. године (око 2.700).²⁸⁹ За све текстове је обезбеђена информација о изворима из којих су преузети, а извршена је и подела текстова по функционалним стиловима (књижевно-уметнички, научни, публицистички, административни и др.).

Корпус се претражује путем регуларних израза једноставном или напредном претрагом. Напредна претрага врши се помоћу CQP (Corpus Query Processor Query Language) упитног језика, а једноставна пружа само основне опције претраживања. Резултат претраге корпуса за дати упит су све ниске које одговарају обрасцу који је описан регуларним изразом. Корисник може да бира део корпуса који ће претраживати, филтрирање резултата претраге по различитим критеријумима и начин исписа корконданци. Могуће је прегледати и шири контекст путем претходног и следећег реда из текста, као и генерисати нове конкорданце.²⁹⁰ „Такође, развијено је природно-језичко сучеље које од корисника не захтева да познају CQL, а омогућава им да једноставно генеришу упит на том језику.“²⁹¹ На слици 22 дат је приказ претраге на природно-језичком сучељу.

The screenshot shows the CQP query interface. At the top, there is a text input field containing the query: `{lemma="jesam"} []{0,2} [pos="V"]`. Below the input field are buttons for '+', '-', 'X', 'Test', 'Formiraj upit', and 'Obriši upit'. Below the input field are three sections for filter criteria:

- Pozicioni atribut 1:** Includes a checkbox for 'Proizvoljni token', a 'Kriterijumi' section with 'Naziv atributa' (lemma) and 'Vrednost atributa' (jesam), and a 'Broj pojavljivanja atributa' (Tačno jednom).
- Pozicioni atribut 2:** Includes a checked checkbox for 'Proizvoljni token', a 'Broj pojavljivanja atributa' section with 'm' (0) and 'n' (2), and a 'Broj pojavljivanja atributa' (Tačno jednom).
- Pozicioni atribut 3:** Includes a checkbox for 'Proizvoljni token', a 'Kriterijumi' section with 'Naziv atributa' (vrsta reči (pos)) and 'Vrednost atributa' (Glagol (V)), and a 'Broj pojavljivanja atributa' (Tačno jednom).

Слика 22 – Природно-језичко сучеље за генерисање CQP-упита²⁹²

²⁸⁹Miloš Utvić. „Izgradnja referentnog korpusa savremenog srskog jezika” (Doktorska disertacija. Filološki fakultet, 2013), 248-259.

²⁹⁰ Корпус савременог српског језика, доступно на: <http://www.korpus.matf.bg.ac.rs>

²⁹¹Милош Утвић. „Листе учестаности Корпуса савременог српског језика.“ *Међународни славистички центар* 43,3(2014):253-254

²⁹²Слика преузета из рада наведеног у претходној фусноти

Корпус СрпКор незаобилазан је код поступака тачног дефинисања и провере појмова за формулисање упита за проналажење и екстракцију информација, па ће у том смислу имати примену у оквиру овог истраживања, у одељцима 6.3.1 и 6.3.2.

Морфолошки речници српског језика

Електронски речник је незаобилазан алат у процесу аутоматске обраде текста. Пошто се користе у области обраде природних језика, потребно је разликовати их од машински читљивих речника. „Електронски речници намењени рачунарској обради текста-корпуса, за разлику од машински читљивог речника који је намењен кориснику рачунара.”²⁹³ Корисник рачунара углавном у речнику тражи значење неке речи или њен превод са једног језика на други док електронски речници имају другу намену и користе се у опсежним и значајним истраживањима језика и креирању језичких алата.

Морфолошке речнике српског језика развили су проф. др Цветана Крстев и проф. др Душко Витас уз помоћ Групе за језичке технологије Математичког факултета. Они су били подстакнути радом Мориса Гроса и морфолошки речници настали су под великим утицајем његове школе. Речници су развијани у оквиру мреже RELEX²⁹⁴ коју су развили Грос и сарадници, а садрже исцрпне описе морфосинтаксичких карактеристика лексике српског језика и снабдевени су семантичким и синтаксним маркерима,²⁹⁵ а базирани су на савременим методама обраде природних језика. Основна јединица морфолошког електронског речника српског језика је лема. Свакој леми у морфолошком речнику придружене су њене морфолошке особине у виду флексије, деривације, рода, броја и сл. исказани одређеним, јасно дефинисаним семантичким и синтаксним маркерима.

За израду морфолошког речника српског језика користи се стандард за електронске морфолошке речнике развијен у лабораторији LADL (Laboratoire d'informatique documentaire et linguistique) на Универзитету у Паризу.²⁹⁶ Код развоја морфолошког речника српског језика, заснованог на поменутом стандарду, требало је

²⁹³Christine Jacquet-Pfau. „Les dictionnaires du français sur cédérom.” *International Journal of Lexicography* 15,1(2002):90

²⁹⁴Université Paris-Est Marne-la-Vallée. LADL – The RELEX Network, доступно на: <http://ladl.univ-mlv.fr/Relex/Relex.html>

²⁹⁵Cvetana Krstev. *Processing of Serbian: Automata, Texts and Electronic Dictionaries*. (Belgrade : Faculty of Philology, 2010), 12-13.

²⁹⁶Blandine Courtois, Max L. Silberztein et al. „Dictionnaires électroniques du français.” *Langue française* 87,1(1990):3-4.

у самом почетку уврстити одређене специфичности српског језика да би речник могао да се примени у анализи и обради најразличитијих текстова:

1. Употреба два алфабета – ћирилица и латиница су у српском језику у равноправној употреби и текстови се појављују на оба алфабета.
2. Правопис заснован на фонологији – употреба различитих варијанти српског језика (екавски и ијекавски изговор) и постојање дублета због те употребе.
3. Богат морфолошки систем који се огледа на нивоима флексије и деривације.
4. Слободан редослед речи у погледу субјекта, предиката, објекта и осталих делова реченице, као и употреба енклитика.²⁹⁷

Електронски речник српског језика Unitex који подржава све побројане особине српског језика развијен је на основу Unitex алата поменуте лабораторије (LADL) и покрива општу лексику, лична и географска имена, лексику везану за енциклопедијско знање, просте и сложене речи српског језика. Речник је настао како на основу традиционалних извора – речника, тако и екстракцијом информација из процесираних текстова. На основу препорука RELEX-а, ресурси за морфолошко процесирање српског језика састоје се од различитих речника – DELAS за просте речи, DELAF за просте речи са флексијама, DELAC за речник сложених речи и DELACF за речник сложених речи са флексијама, као и од морфолошких граматика и локалних граматика.

Одредници у DELAS и DELAC речнику придружена је ознака врсте речи (једно или више великих слова), веза са подређеним облицима (нумеричка или алфанумеричка ознака која заједно са врстом речи омогућава аутоматско генерисање свих подређених облика за DELAF речник), синтаксни, семантички, употребни, дијалекатски и сл. маркери (слободни маркери, алфанумеричке ознаке). Одредници у DELAF и DELACF речнику придружен је канонски облик (лема), ознака врсте речи која је преузета од леме, затим синтаксни, семантички, дијалекатски и сл. маркери такође преузети од леме, као и скуп кодова који су везани за вредност граматичких категорија које се односе на облик (једномесни кодови састављени од великог или малог слова или цифре).

²⁹⁷Duško Vitas and Cvetana Krstev. „Derivational Morphology in an E-Dictionary of Serbian.” In: Zygmunt Vetulani (ed.) *Proceedings of 2nd Language & Technology Conference*, April 21-23, 2005, Poznań, Poland, pp.139-143 (Poznań: Wydawnictwo Poznańskie, 2005):140.

Листа морфолошких категорија и кодова у електронском речнику облика српског језика представљена је у Табели 2, а неки од маркера придружених лемама у српском DELAS речнику приказани су у Табели 3.²⁹⁸

Табела 2 – Листа морфолошких категорија и кодова у морфолошком речнику српског језика

Врста речи	Категорија	Код	Врста речи	Категорија	Код	
именице – N	Род	m – мушки	придев– А	род		
		f – женски		падеж		
		n – средњи		број		
	падеж	1– номинатив			аниматност	
		2 – генитив			степен поређења	a – позитив
		3 – датив				b – компаратив
		4 – акузатив				c – суперлатив
		5 – вокатив			одређеност	d – одређен
		6– инструментал				k– неодређен
		7 – локатив			глаголи– V	облик
	број	s – једнина				P – презент
		p – множина				A – аорист
	аниматност	v –живо				I – имперфекат
		q – неживо				Y – императив
g – без значаја		G – актив				
предлози– PREP	/	/			T– пасив	
конјукција – CONJ	/	/			F – футур	
речце– PAR	/	/			S – гл. прилог садашњи	
Узвици – INT	/	/			X – гл. прилог прошли	
Бројеви – NUM	падеж			лице	x – прво	
	број				y – друго	

²⁹⁸Cvetana Krstev, Duško Vitas and Gordana Pavlović-Lažetić. „Resources and methods in the morphosyntactic processing of Serbo-Croatian.” In: Gerhild Zybatow et al. (eds.) *Formal Description of Slavic Languages: The Fifth Conference*, Leipzig 2003, pp. 3-17 (Frankfurt am Main, Peter Lang 2008).

заменеце – PRO	род				z – треће
	аниматност			негација	h – негативна клитика
	падеж				i – позитивна клитика
	број			број	
	род				
	аниматност				
	наглашеност	i – наглашено			

Како морфолошке категорије и маркери који су придружени лемама изгледају у морфолошком речнику српског језика, показаћемо и на примерима за просте и сложене речи:

а) проста реч

publici, publika.N:fs3q

(именица женског рода, једнина, у дативу, аниматност неживо)

б) сложена реч

baze podataka, baza podataka.N+Comp:w2qf

(именица, сложена реч, женски род, паукал, генитив, аниматност неживо)

Укупан број одредница у електронском речнику српског језика је око 125.000, а број облика је преко 4 милиона. Број одредница сложених речи у DELAC/DELACF речнику је 2008. године био приближно 12.500, али се тај број повећава захваљујући пројектима који се баве искључиво богаћењем морфолошког речника сложених речи. Један од таквих пројеката јесте и ово истраживање.

Табела 3 – Приказ маркера придружених лемама у српском DELAS речнику

Категорија	Таг (ознака)	Одговара	Објашњење	Пример
Синтакса	+p2	Предлозима	Именичка фраза у генитиву	bez,PREP+p2
	+Ref	Глаголима	Рефлексивност	dičiti,V551+Imperf+It+Ref
	+MG	Именицама	Природан род, мушки	budala,N601+Hum+MG+FG
Деривација	+VN	Именицама	Глаголска именица	kiseljenje,N300+VN
	+Adj	Прилозима	Изведени од придева	fanatično,ADV+Adj

	+Der=OvatiIrtati	Глаголима, именицама, прилозима	Деривациона гнезда	dezinfikovati,V18+...+Der=OvatiIrtati
Семантика	+Col	Прилозима	Боје	zelenkastosiv,A6+Col
	+Hum	Именицама	Људи	ljubavnica,N601+Hum
	+Mat	Придевима	Материјал	kožnat,A6+Mat
Дијалекти	+Ek	Свим врстама речи	Екавски	nedelja,N600+Ek
	+Ik	Свим врстамаречи	Ијекавски	nedjelja,N600+Ik

За квалитетно претраживање комплетног текста није довољно спровести само добро оптичко препознавање карактера. Имајући у виду сложен систем флексије у српском језику, није довољно претраживање само према основним облицима речи (нпр. номинативу или инфинитиву) јер ће тако у резултатима претраживања изостати сви други облици речи који се појављују у тексту, а који су у директној вези са постављеним упитом. Морфолошки речници помажу да претраживање да квалитетнији одзив и да се пронађу све информације које су у складу са постављеним упитом.

Применом свих поменутих морфолошких речника српског језика на корпусу текстова предвиђених овим истраживањем, а коришћењем алата Unitex²⁹⁹ који омогућава изградњу, проверу и примену електронског речника, као и претрагу путем регуларних израза (токена³⁰⁰, конкатенације³⁰¹, уније регуларних израза³⁰² и

²⁹⁹Sébastien Paumier. *Unitex 3.1 beta : user manual* (Paris : Université Paris-Est Marne-la-Vallée, 2003-2014), <http://www-igm.univ-mlv.fr/~unitex/UnitexManual3.1.pdf> (приступљено 25.09.2014)

³⁰⁰ Корпусна реч као низ карактера између два узастопна сепаратора где је сепаратор скуп неалфанумеричких карактера

³⁰¹Надовезивање токена

³⁰²Регуларни израз је текст комбинован са специјалним карактерима (који се називају метакарактери), а који представља објекте који се не могу непосредно унети са тастатуре као што су „крај линије ” или „ма који број” или „три ма која велика слова једно до другог.”” – Милош Утвић. „Регуларни изрази.” *Свет речи* 15-16(2003):66.

Клинијевог затворења регуларног израза³⁰³) и графова, биће могуће утврдити које се просте и сложене речи налазе у постојећим морфолошким речницима српског језика, а такође утврдити и листу речи које су остале непрепознате после аутоматског процесирања текста.

Прегледом и анализом те листе, а консултовањем конкорданци у СрпКор у одељку 6.3.2 ћемо дати предлоге за допуну и измену постојећих речника, посебно у домену сложених речи чиме ће се значајно допринети порасту броја одредница у DELAC/DELACF речнику, са једне стране, а тиме и бољем проналажењу информација у процесираним текстовима дигиталних библиотека, са друге стране.

Српску wordnet

Српски wordnet заснован је на структури америчког wordnet-а за енглески језик тзв. Princeton wordnet-а, односно Пристонског wordnet-а.³⁰⁴ Пристонски wordnet развио је Џорџ Милер (George Miller) са својим тимом у оквиру пројеката везаних за психолингвистику почев од 1985. године. Принстонски wordnet конципиран је тако да је сваки концепт у њему „мотивисан скупом синонима у енглеском језику“.³⁰⁵ Wordnet је врло велика лексичка база података која је организована преко чворова и релација између тих чворова творећи тако семантичку мрежу. Ови чворови се у wordnet-у називају синсетови (од енгл. synset или synonymous set) и представљају скупове речи које у неком контексту имају исто (синонимно) значење.

Данас се изградња wordnet-а креће у правцу глобалног wordnet-а који не би понудио само компаративну језичку анализу већ и управљање знањем, управљање садржајем, екстракцију и проналажење информација, машинско превођење базиране на концепту вишејезичности.³⁰⁶ Уопште гледано, глобални wordnet је корисна компонента у аутоматској обради текста.

³⁰³ Клинијево затворење или Клинијева звезда (*) – симбол који је увео Стивен Клини, а примењује се на скуповима ниски, симбола или карактера приликом претраживања регуларним изразима

³⁰⁴ Cvetana Krstev, Gordana Pavlović-Lažetić and Ivan Obradović. „Using textual and lexical resources in developing Serbian Wordnet.” *Romanian Journal of Information Science and Technology* 7,1-2(2004):147-161.

³⁰⁵ Cvetana Krstev. „Specifični koncepti Balkana u semantičkoj mreži Wordnet.” U: Ljiljana Subotić et al (ur.) *Zbornik radova „Susreti kultura“*, Novi Sad, decembar 2004, pp. 275-285. (Novi Sad: Univerzitet, Filozofski fakultet, 2006):275

³⁰⁶ Cvetana Krstev, Svetla Koeva and Duško Vitas. “Towards the Global Wordnet.” In: *Conference Abstracts of the First International Conference of Digital Humanities Organizations (ADHO) Digital Humanities 2006*, Paris-Sorbonne, 5-9 July 2006, p. 114-117 (Paris: Université Paris-Sorbonne, 2006):114.

Српски wordnet као семантичка база развијен је у оквиру пројекта BalkaNet који се бави развојем wordnet-а као вишејезичке базе података за следеће балканске језике: грчки, турски, бугарски, румунски, српски, а и за чешки језик који је придодат овом пројекту да би се остварио континуитет са моделом EuroWordnet-а. Пројекат EuroWordnet који је у концепт wordnet-а увео вишејезичност обухватао је следеће језике: холандски, италијански, шпански, немачки, француски, чешки и естонски.

У изградњи wordnet-а у оквиру пројекта BalkaNet одлучено је да се задржи структура коју је развио Принстонски wordnet без обзира на тешкоће које су се јављале приликом преузимања концепата и дефинисања синсетова за балканске језике. Један од разлога јесте идеја о вишејезичним међусобно поравнатим концептима нарочито значајним код вишејезичког претраживања и других вишејезичних примена. Свакако, због специфичности везаних за историјске, друштвене, географске, породичне и друге прилике, свака балканска земља учесница пројекта је у wordnet за свој језик уврстила и неке нове концепте који нису садржани у Принстонском wordnet-у.³⁰⁷

Како се аутоматска изградња wordnet-а најчешће заснива на коришћењу вишејезичких лексичких и, посебно, текстуалних ресурса у дигиталном облику³⁰⁸ у оквиру ових истраживања биће могуће допунити и доградити српски wordnet новим концептима. Концепти су, у првом реду, везани за област библиотекарства и информатике јер је корпус текстова на којима се врши истраживање управо из ове области. У неким случајевима, на основу резултата аутоматске анализе, биће потребно извршити и ручну допуну термина. Валидација термина ће се вршити на примерима из СрпКор-а као и провером у морфолошком речнику српског језика.

Терминолошки речник из библиотекарства и информатике

Да би се код екстракције термина из одређене области добили добри резултати, неопходно је на корпус текстова за анализу применити неки терминолошки речник. Пошто ћемо анализирати корпус текстова из библиотекарства и информатике, као основа за овај речник послужиће база двојезичног српско-енглеског и енглеско-српског „Библиотекарског терминолошког речника“ ауторки Љиљане Ковачевић,

³⁰⁷ Cvetana Krstev. „Specificni koncepti Balkana u semantičkoj mreži Wordnet.” U: Ljiljana Subotić et al (ur.) *Zbornik radova „Susreti kultura“*, Novi Sad, decembar 2004, ured., pp. 275-285 (Novi Sad: Univerzitet, Filozofski fakultet, 2006).

³⁰⁸ Цветана Крстев [и др.]. „Кооперативни рад на доградњи српског Wordneta.” *Инфотека* 9,1-2(2008): 57-75.

Весне Ињац и Добриле Бегенишић, у издању Народне библиотеке Србије 2004. године.³⁰⁹ Добра страна овог речника је што осим штампаног издања, постоји издање на CD-у из 2002, као и онлајн верзија речника у коме је могуће претраживати термине на српском и енглеском језику.³¹⁰ Речник је наставак и надградња међународног пројекта „Multilingual dictionary of library terminology“ који је покренут 1990. са циљем да представи терминологију из библиотекарства и информатике на 16 европских језика. Координатор за српски језик, Народна библиотека Србије, пројекту је приступила 2000. године. Онлајн верзија српско-енглеског/енглеско-српског речника коју смо користили за потребе овог истраживања садржи 23.400 термина, од којих је 12.100 на српском, а 11.300 на енглеском језику.

Нова верзија овог речника је српско-енглеско-немачки термилошки речник који носи назив „Речник библиотекарства и информационих наука“ и израђен је 2014. године.³¹¹ Речник није публикован као штампано издање, већ само као електронски извор онлајн и садржи око 40.000 одредница (14.000 на српском – око 2.000 нових термина у односу на описану верзију, 12.400 на енглеском и 14.000 на немачком).³¹²

Садржај „Библиотекарског термилошког речника“ чине следеће области:

- библиотекарство: библиотечке колекције, библиографски опис и садржинска анализа, класификација и индексирање, тезауруси, чување, претраживање и дисеминација докумената и информација, рад са корисницима, библиотечко особље, опрема и простор;
- рубне области: архивистика, библиографија, графички дизајн, историја писма и књиге, издаваштво и књижевство, штампарство и репрографија, заштита и конзервација фондова, лингвистика и терминологија, менаџмент и маркетинг;
- рачунарство и комуникационе науке: електронски извори, рачунарски производи и услуге, информациони системи, телекомуникације, интернет,

³⁰⁹Љубазношћу ауторки, за потребе израде ове дисертације, добијена је на коришћење база података са речима које су саставни део овог речника и њиховим дескрипторима

³¹⁰ Библиотекарски термилошки речник, доступно на: <http://btr.nb.rs/default.htm>

³¹¹Љиљана Ковачевић, Добрила Бегенишић, Весна Ињац-Малбаша. *Речник библиотекарства и информационих наука: српско-енглеско-немачки*. (Београд: Народна библиотека Србије, 2014), доступно на: <http://rbi.nb.rs/sr/home.html>

³¹²Сматрамо да је неопходно констатовати да се у изради допуне „Библиотекарског термилошког речника“ не користе језичке технологије и алати који су развијени за српски језик, а који би у многоме олакшали рад на овако обимном послу.

масовне комуникације, електронска обрада података, рачуарска опрема, рачуарски програми (софтвер);

- правни аспект информација и документације: ауторско право, заштита интелектуалне својине, заштита података, право на информисање.³¹³

Имајући у виду овако велики број области које речник покрива, ауторке су свакој одредници у речнику доделиле дескриптор у виду скраћенице. Једна одредница може имати један или више дескриптора. Попис свих дескриптора са објашњењима дат је у табели 4.

Табела 4 – Дескриптори речи са објашњењима у „Библиотекарском терминолошком речнику“

Дескриптор	Објашњење	Дескриптор	Објашњење
архи	архивистика	кор	корисници
арт	уметност	лег	право, закони, прописи
бибг	библиографија	линг	лингвистика
библ	библиотекарство	ману	рукописи
цирк	циркулација	мнг	руковођење, менаџмент
доку	документалистика	музе	музеологија
еду	образовање	музи	музикалије
елр	електронски извори	наб	набавка
еоп	електронска обрада података	нкг	некњижна грађа
фин	финансије	огл	оглашавање
фонд	фонд	опр	опрема
ген	генерално, опште	орг	организација
икон	илустрације, украси	пап	папир
инд	индексирање	перс	особе, личности
инф	информатика	пов	повез
интер	интернет	пр	односи са јавношћу
карт	картографија	принт	штампарство
кат	каталогизација	публ	издаваштво
клас	класификација	репр	репрографија
кнм	управљање знањем	сцри	писање
кол	колекција, збирка	сسر	претраживање
ком	комуникологија	тез	тезауруси
конз	конзервација	згра	зграде, објекти

Да би термини из овог речника постали саставни део морфолошког речника српског језика, свакој класификованој категорији термина, на основу дескриптора,

³¹³ Наведено према „Библиотекарски терминолошки речник.“ О речнику, доступно на: <http://btr.nb.rs/recnik.htm>

додељени су семантички маркери.³¹⁴ Тако су сви термини из овог речника добили почетни маркер ВІ (библиотекарство и информатика) који је допуњен маркером одређеног дескриптора (нпр. ВІelr – електронски извори, ВІeop – електронска обрада података, ВІind – индексирање, ВІinf – информатика, ВІinter – интернет, ВІling – лингвистика, ВІtez – тезауруси...). Доделом овако структурираних семантичких маркера задржана је веза са изворним речником из кога су термини преузети, а јасно је назначена припадност одређеном поддомену домена библиотекарство и информатика. Како поједини термини речника могу имати један или више дескриптора, тако и у електронском морфолошком речнику српског језика, један термин може имати један или више еквивалентних семантичких маркера (нпр. *аутоматско превођење* – ВІeop + ВІling; *дигитална библиотека* – ВІbibl + ВІeop + ВІorg).

Табеле са речима из овог речника и њихови дескриптори били су од великог значаја за допуну лексичких ресурса, српских и двојезичних, који су потом допринели овом истраживању. У даљем тексту биће објашњена њихова употреба првенствено код рада са алатом *Библиша*.

За потребе допуне постојећих лексичких ресурса, кренули смо од класа речи које покривају област информатике и библиотечке информатике и према понуђеним табелама извршили валидацију и одабир термина из следећих класа: ВІelr, ВІind, ВІinf, ВІinter, ВІklas, ВІling, ВІssr, ВІtez, ВІeop. Одабрали смо 2.323 термина којима је започета израда електронског морфолошког термилошког речника из библиотекарства и информатике.³¹⁵ У следећим корацима смо за допуну речника користили методе аутоматске екстракције термина из доменског корпуса што ће детаљно бити објашњено у одељку 6.3.1.

На основу „Библиотекарског термилошког речника“ као класичног и почетног термилошког извора и његове двојезичне српско-енглеске базе, а помоћу раније наведених морфолошких речника српског језика и система Unitex, у истраживању ћемо анализирати могућност допуне специфичног термилошког речника српског језика из области библиотекарства и информатике, као и потенцијалну допуну општег речника српског језика оним речима које су из домена

³¹⁴ Семантичке маркере доделила је проф. др Ранка Станковић

³¹⁵ Aleksandra Trtovac and Jelena Andonovski. „Enrichment of Morphological Dictionary of MWUs.” In: Pavlović-Lažetić, Gordana et al. (eds.), *Natural Language Processing for Serbian*, pp. 27-40. (Belgrade: University of Belgrade, Faculty of Mathematics, 2014):32-35

библиотекарства и информатике прешле у општу употребу код говорника српског језика.

Осим допуне морфолошког речника српског језика простим речима (DELAS/DELAF речници) очекујемо у значајној мери и пораст броја сложених речи (DELAC/DELACF речници). Показаћемо и како се применом базе речи овог речника може поспешити проналажење и екстракција информација у дигиталним ресурсима из области библиотекарства и информатике.

Корпуси поравнатих текстова

СрпКор, осим корпуса текстова на српском језику, садржи и француско-српски и енглеско-српски паралелни корпус. Развој паралелних корпуса који укључују српски као један језик започео је учешћем Математичког факултета Универзитета у Београду у пројекту TELRI³¹⁶ који је допринео да српски језик буде један од језика за које је извршена паралелизација текста Платонове „Државе“ (паралелни текст на 16 језика – бугарски, енглески, летонски, литвански, мађарски, немачки, пољски, румунски, руски, словачки, словеначки, српски, фински, француски, хрватски, чешки) и у пројекту MULTEX-East где је паралелизован текст Орвелове „1984“ (паралелни текст на 12 језика – бугарски, енглески, естонски, литвански, мађарски, резизијански³¹⁷, румунски, руски, словеначки, српски, хрватски, чешки). Текст на сваком појединачном језику упарен је са текстом на енглеском језику.

Након завршених наведених пројеката наставило се са радом нарочито на француско-српском паралелном корпусу књижевних (Волтеров „Кандид“, Вернов „Пут око света за осамдесет дана“) и новинских текстова („Le monde diplomatique“) где је француски изворни језик, а српски је циљни језик (језик превода).³¹⁸

Рад на Верновом делу „Пут око света за осамдесет дана“ остварен је у оквиру независног пројекта, а у сарадњи Математичког и Филолошког факултета и укључује 18 језика за које је урађен паралелни корпус (албански, бугарски, грчки, енглески, италијански, кинески, мађарски, македонски, немачки, пољски, португалски, румунски, руски, словеначки, српски, француски, хрватски, шпански).³¹⁹

³¹⁶ TELRI – Trans European Language Resources Infrastructure, доступно на: <http://telri.nytud.hu/>

³¹⁷ Резизијански језик је аутономно наречје словеначког језика којим се говори у области Резија у Италији

³¹⁸ Cvetana Krstev. *Processing of Serbian: Automata, Texts and Electronic Dictionaries*. (Belgrade: Faculty of Philology, 2008), 159.

³¹⁹ META-SHARE. Multilingual Edition of Verne's Novel "Around the World in 80 Days", <http://metashare.nlp.ipipan.waw.pl/metashare/repository/browse/multilingual-edition-of-vernes-novel-around->

Енглеско-српски паралелни корпуси у оквиру СрпКор-а налазе се у оквиру два поткорпуса: SELFЕН³²⁰ састављен од текстова из домена финансија, здравља, закона и образовања и BALKANTIMES који садржи вести са сајта Southeast European Times (SETimes) са информацијама о Југоисточној Европи (текуће вести, економија, дипломатија, филм, туризам, спорт, наука) на седам балканских језика (албански, босански, македонски, српски, руски, украјински, хрватски) и упоредо на енглеском језику.³²¹

Поравнати текстови, као што се види из наведених примера, третирају двојезичне или вишејезичне паралелизоване корпуре. Основни елемент двојезичног или вишејезичног паралелизованог корпуса је битекст или паралелизовани текст. Битекст подразумева текст и његов превод између којих је успоставља логичка експлицитна веза.³²² Изворни (основни) текст и његов превод (на циљни језик), односно семантички еквивалент, поравнавају се (паралелизују) до различитих нивоа – параграфа, реченица, фраза или речи. Потпуно паралелизовање текстова остварено је када су јединице текста једног језика у потпуности поравнате са јединицама текста другог језика.³²³

Језички алати којима се формирају паралелизовани текстови (битекстови) које користи Група за језичке технологије Универзитета у Београду поравнавају текстове на нивоу одељака, пасуса и сегмената (најчешће реченица). Сегменти могу бити структурирани на више начина:

- сегмент изворног језика је једна реченица и сегмент циљног језика је једна реченица;
- сегмент изворног језика је једна реченица, а сегмент циљног језика су две или више реченица (јавља се и обрнут случај);

the-world-in-80-days/438d369a8bdf11e2b44a00151714459236eaa2be46a54847812a1f210d9f378e/
(приступљено 30.10.2015)

³²⁰SELFЕН – Serbian-English Law Finance Education and Health

³²¹SETimes. О нама,
http://www.setimes.com/cocoon/setimes/xhtml/sr_Latn/document/setimes/footer/about/about (приступљено 22.10.2014)

³²²Miloš Utvić. „Izgradnja referentnog korpusa savremenog srpskog jezika.” (Doktorska disertacija. Filološki fakultet, 2013), 30.

³²³Зоран Ристовић. „Од корпуса до учионице – примена паралелизованих текстова у настави енглеског језика у основној школи.” *Инфотека* 13, 2(2012):52-66.

- сегмент изворног језика је једна реченица, а сегмент циљног језика је део реченице (и обрнуто);
- сегмент изворног језика је једна реченица, али не постоји одговарајући сегмент циљног језика (и обрнуто);
- сегменти изворног и циљног језика састоје се од две или више реченица које нису у идентичном редоследу у изворном и циљном тексту.

Све су ово проблеми који се могу јавити код поравнавања до нивоа реченице, нарочито код књижевних текстова, па је јасно уочљиво да се „сравњивање“ текста до нивоа речи или фразе још тешко постиже.

Међутим, без обзира на све тешкоће које се јављају приликом паралелизације текстова, двојезични и вишејезични паралелни корпуси имају значајну улогу у обради природних језика, екстракцији терминологије из одређене области, машинском превођењу, контрастивној лингвистици, учењу страних језика, као и у лексикологији.

Поравнати корпуси се често користе и код језика који су веома слични, па се из тог разлога ретко врши међусобно превођење (српски, хрватски и босански, чешки и словачки).³²⁴ У том случају упарују се преводи дела са страних језика на ове језике што се и види из примера који су наведени (Платонова „Држава“, Волтеров „Кандид“, Вернов „Пут око света за осамдесет дана“, Орвелова „1984“ и сл).

За проналажење информација у дигиталним изворима и дигиталним библиотекама корпуси паралелних текстова су изузетно значајни јер омогућавају претраживање на различитим језицима које резултира и релевантним погоцима на другим језицима, а доприноси и критеријуму вишејезичности, као и бољој видљивости нашег језика у виртуелном свету.

Ради допуне електронског термилошког речника из области библиотекарства и информатике, у овом истраживању анализираће се корпус поравнатих текстова из ове области, на српском и енглеском језику, у првом реду корпус поравнатих српских и енглеских електронских текстова у двојезичном часопису „Инфотека“ (на енг. „Infotheca“).³²⁵

³²⁴ Cvetana Krstev, Duško Vitas, „An Aligned English-Serbian Corpus”, In: N. Tomović & J. Vujić (eds.) *ELLSIIR Proceedings (English Language and Literature Studies: Image, Identity, Reality)*, Vol. 1, Belgrade, 4-6 December 2009, pp. 495-508. (Belgrade: University, Faculty of Philology, 2011)

³²⁵ Часопис излази од 2000. године, а од 2004. године као двојезично издање. Корпус паралелних текстова третира све текстове који се јављају упоредо на српском и енглеском језику од 2004. до 2014. године. Издавач часописа је до 2013. била Заједница библиотека универзитета у Србији, а од 2014.

Лексички алат Библиша

Неки од наведених лексичких ресурса садржани су у једном алату који је развила Група за језичке технологије Универзитета у Београду. *Библиша*³²⁶ је веб апликација помоћу које се претражује терминима у текстовима чланака објављеним у електронским часописима који излазе на два језика (за сада су то српски и енглески), односно уопште поравнати текстови (нпр. предлози и извештаји пројеката). Овим алатом могуће је проналажење информација на српском и енглеском језику, постављањем упита простим, али и вишечланим речима.³²⁷ Упити могу бити проширени, како семантички тако и морфолошки што је изузетно значајно код језика богате флексије, какав је српски језик.³²⁸

Алат је изузетно значајан и због тога што обједињује проналажење информација и преко дескриптора метаподатака и преко дескриптора садржаја. Валидација ове апликације је извршена на корпусу поравнатих текстова поменутог часописа „Инфотека.” Реч је о 74 чланка објављена у периоду од 2007. до 2014. из дигиталне библиотеке овог часописа³²⁹ који су упоредо на српском и енглеском језику, а тестирања су вршена на ТМХ³³⁰ колекцији докумената.

Лексички ресурси

Лексички ресурси који су се примењивали у претраживањима овог корпуса су морфолошки речници српског језика, српски и енглески wordnet и база двојезичног „Библиотекарског термилошког речника.”

издавачи су и Универзитетска библиотека „Светозар Марковић“ и Филолошки факултет у Београду. Часопис Инфотека, доступно на: <http://infoteka.bg.ac.rs>.

³²⁶ Bibliša: Aligned Collection Search Tool, доступно на: <http://hlt.rgf.bg.ac.rs/Bibliša/>

³²⁷ Под вишечланом речи (полилексемом, лексичким изразом) подразумевамо српски еквивалент за устаљени термин у англосаксонској литератури Multi-Word Units (MWU) – „ниска речи са јединственим значењем” – Norbert Schmitt. *Vocabulary in language teaching* (Cambridge: Cambridge University Press, 2000), 96

³²⁸ Ranka Stanković, Cvetana Krstev, Ivan Obradović, Aleksandra Trtovac and Miloš Utvić. „A tool for enhanced search of multilingual digital libraries of e-journals.” In: Nicoletta Calzolari [et al.] (eds.) *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12)*, Istanbul, May 23-25, 2012, pp. 1710-1717 (Istanbul: European Language Resources Association (ELRA), 2012), http://www.lrec-conf.org/proceedings/lrec2012/pdf/375_Paper.pdf (приступљено 07.12.2015)

³²⁹ Bibliša. Journal Selection, <http://hlt.rgf.bg.ac.rs/Bibliša/Documentation.aspx> (приступљено 07.12.2015)

³³⁰ ТМХ – Translation Memory eXchange – Преводилачка меморија за размену, доступно на: <http://www.gala-global.org/oscarStandards/tmx/tmx14b.html>

У првој фази коришћен је само корпус паралелних текстова часописа „Инфотека“ на којима је и вршена валидација алата. База *Библиша* у овом тренутку садржи осим радова из „Инфотеке“ и паралелизоване чланке из следећих часописа: „Подземни радови“ („Underground Mining Engineering“), „Архитектура и урбанизам: часопис за архитектуру, урбанизам и просторно планирање“ („Architecture and Urbanism: Journal of architecture, urbanism and spatial planning“) и „Стоматолошки гласник Србије“ („Serbian Dental Journal“), као и предлоге и извештаје пројеката. Сви ови часописи излазе двојезично и текстови су упарени до нивоа реченице.

Проналажење и екстракција информација у овим паралелизованим корпусима текстова врши се данас путем различитих лексичких ресурса (морфолошки речници српског језика, српски и енглески wordnet, база двојезичног „Библиотекарског термилошког речника“, RudOnto – двојезични речник из рударства, GeolISS – двојезични речник из геологије, а у изради је и Stomatolog – двојезични речник из стоматологије).

У овом истраживању базираћемо се на претрази корпуса текстова из часописа „Инфотека“ јер је тај корпус основ за изградњу и допуну електронског термилошког речника из библиотекарства и информатике.

Технологије

Претрага терминима из наведених ресурса подржана је технологијама које су уграђене у апликацију *Библиша*:

1. Leximir³³¹ – лексички алат за руковање лексичким ресурсима и за проширење упита који садржи неколико компоненти за извршавање различитих функција:
 - конверзија омогућава трансформисање ресурса из једног формата у други (углавном је реч о конвертовању изједне кодне шеме у другу текстова, речника, графова и регуларних израза);
 - подсистем за одржавање морфолошких речника – пружа могућност претраживања и издвајања лема на основу поређења са задатом ниском карактера, врстом речи, кодом флективне класе,³³² синтаксним и семантичким ознакама; обезбеђује се веза са графовима који описују флективна својства

³³¹ Leximir, доступно на: <http://korpus.matf.bg.ac.rs/soft/LeXimir.html>

³³² Кодови флективних класа идентификују флективне графове помоћу којих се генеришу сви флективни облици дате леме.

леме, а едитор леме омогућава формирање нове леме из почетка или копирање сличне леме која се затим модификује;

- развој и унапређење wordneta подразумева подршку раду са одређеним wordnet-ом као и синхронизацију рада два wordnet-a који се остварује преко јединственог идентификатора ILI (Inter-Lingual Identifier);
 - подсистем за интеракцију система електронских речника и семантичке мреже Wordnet омогућава размену информација између wordnet-a и морфолошких речника – морфосинтасичке информације из речника се придружују литералима у синсетовима, а семантичке информације из синсетова лемама у речницима;
 - изградња, коришћење паралелизованих текстова и конверзија из TEI формата у друге формате, посебно у TMX;
 - подсистем за генерисање класа вишечланих речи на основу морфолошке анализе, система творбених правила и описа морфолошких особина српског језика уз коришћење графова³³³;
 - подсистем за предикцију класе вишечланих речи и генерисање речничке одреднице целе леме за задату листу термина;
 - подсистем за предикцију флективне класе (простих) речи заснован на правилима која користе метаподатке о флективним класама и статистику суфикса постојећих речничких одредница;
 - подсистем за издвајање и статистичку анализу именованих ентитета;
 - подсистем за аутоматско издвајање термина из текста и генерисање исправне леме и целе речничке одреднице.
2. VebRanka³³⁴ – алат је надградња алата Leximig и намењен је проширивању упита са циљем да омогући повећање одзивакод претраживања веба коришћењем језичких ресурса за српски језик³³⁵; VebRanka користи сервис

³³³Cvetana Krstev, Ranka Stanković, Duško Vitas, Ivan Obradović, „WS4LR – a Workstation for Lexical Resources”. In: *Proceedings of the Fifth Interantional Conference on Language Resources and Evaluation*, Genoa, Italy, May 2006, pp. 1692-1697, 2006:1693-1695, <http://www.lrec-conf.org/proceedings/lrec2006/> (приступљено 30.10.2015)

³³⁴ VebRanka, доступно на: <http://hlt.rgf.bg.ac.rs/vebranka/>

³³⁵Cvetana Krstev, Ranka Stanković, Duško Vitas, Ivan Obradović. „The Usage of Various Lexical Resources and Tools to Improve the Performance of Web Search Engines.” In: *Proceedings of the Sixth Interantional Conference on Language Resources and Evaluation (LREC'08)*, Marrakech, Morocco, 28-30 May 2008, pp. 219-224. European Language Resources Association (ELRA), 2008:220-221, <http://www.lrec-conf.org/proceedings/lrec2008/> (приступљено 30.10.2015)

VebRan³³⁶ у оквиру ког су јавно доступне различите лингвистичке операције за семантичко и морфолошко проширење упита, као и проширење упита на још један језик зависно од расположивих ресурса. Када је у питању семантичко проширење, у зависности од спецификованих захтева корисника позива се оговарајућа функција, рецимо `getSinonimi` уколико је потребно проширити упит само синонимима или `getSinonimiRelSSWN` уколико је потребно проширење упита литералима надређених и подређених концепата, уз могућност проширења и на други језик који се проследи као параметар. Операције за проширење упита веб сервиса VebRan примењенесу у алату *Библиша*.

3. TMX³³⁷ колекција докумената – састоји се од чланака у формату паралелизованих текстова TMX чија је намена једноставна размена података између различитих преводачких алата; један TMX документ се састоји од заглавља и тела документа, у заглављу се налазе метаподаци о паралелизованом тексту документа, а у телу документа је скуп јединица превођења састављен од две или више семантички еквивалентне јединице превођења (на првом месту се јавља изворни језик, а на другом циљни језик или језици). Сам TMX је XML документ.
4. MarkLogic³³⁸ XML база података – база подржава могућност свођења речи на коренску реч у различитим језицима, укључујући енглески језик; за српски језик ова технологија није предвиђена; морфолошки облици у упитима кључним речима на енглеском језику зависе од могућности које пружа MarkLogic, док за српски језик *Библиша* мора да прошири иницијални упит свим морфолошким облицима кључне речи користећи лексичке ресурсе за српски језик.

Објаснићемо како алат *Библиша* ради коришћењем наведених технологија. После постављеног упита кључном речју, *Библиша* позива сервис VebRan да би се пронашли одговарајући термини, синоними и њихови преводи, односно упит проширио семантички и на други језик. Када корисник сузи понуђено проширење полазне кључне речи на одговарајућу листу термина, систем проширује морфолошки упит, и

³³⁶VebRan, доступно на: <http://hlt.rgf.bg.ac.rs/vebran/service.aspx>

³³⁷TMX – Translation Memory eXchange – Преводачка меморија за размену

³³⁸ MarkLogic, доступно на: <http://www.marklogic.com/>

затим трансформише упит у XQuery³³⁹ облику који се затим користи за претрагу ТМХ колекције докумената. Упит у XQuery облику прослеђује колекцији ТМХ паралелних текстова које складишти MarkLogic систем намењен за управљање XML базама података.

Навешћемо пример на коме се виде семантичко и морфолошко проширење упита за термин *књига* у коме се овај термин семантички проширује термином *публикација* уз морфолошко проширење, као и проширење на други језик (енглески) терминима *book* и *publication*:

```
http://biba.rgf.bg.ac.rs:8005/wsQ4.xqy?qJCID=0&request=<request>
<query xml:lang='sr'>knjiga</query>OR<query xml:lang='sr'>knjigama</query>OR
<query xml:lang='sr'>knjige</query>OR<query xml:lang='sr'>knjigo</query>OR
<query xml:lang='sr'>knjigom</query>OR<query xml:lang='sr'>knjigu</query>OR
<query xml:lang='sr'>knjizi</query>OR<query xml:lang='sr'>publikacija</query>OR
<query xml:lang='sr'>publikacijama</query>OR<query xml:lang='sr'>publikacije</query>OR
<query xml:lang='sr'>publikaciji</query>OR<query xml:lang='sr'>publikacijo</query>OR
<query xml:lang='sr'>publikacijom</query>OR<query xml:lang='sr'>publikaciju</query>OR
<query xml:lang='en'>book</query>OR<query xml:lang='en'>publication</query></request>
```

За енглески језик морфолошко проширење није потребно јер су морфолошке особине енглеског језика, као и још неких језика са великим бројем говорника (немачки, француски, руски, италијански, шпански, арапски, кинески, јапански итд.³⁴⁰), уграђене у MarkLogic.

Делови упита и ТМХ докумената који су означени атрибутом `xml:lang='sr'` се сравњују. Затим се проналазе конкорданце у којима се јавља неки од облика речи из упита, а атрибутом `xml:lang='sr'` се филтрирају језик упита и језик документа.

³³⁹XQuery је језик за налажење и екстракцију елемената и атрибута из XML докумената, XQuery 3.0 – an XML Query Language, <http://www.w3.org/TR/xquery-30/> (приступљено 7.12.2015)

³⁴⁰MarcLogic. „Language Support“, http://www.marklogic.com/resources/marklogic-language-support/resource_download/datasheets/ (приступљено 21.12.2015)

BIBLIŠA: ALIGNED COLLECTION SEARCH TOOL

Home Journal selection Metadata search Metadata browse MarkLogic search Biblimir Help About

WELCOME TO JOURNALS ALIGNED COLLECTION SEARCH TOOL!

Keyword Journal(s)


Synonyms	sr	en
<input type="checkbox"/> WordNet...		
<input checked="" type="checkbox"/> Dictionary of Librarianship ...	baza podataka	database
<input type="checkbox"/> Biblimir ...		
<input type="checkbox"/> GeolISSTerm...		
<input type="checkbox"/> RudOnto ...		

Include Hypernyms Include Hyponyms

Serbian query English query

Слика 23 – Пример постављеног упита у алату Библиша

Наведене и описане технологије које потпомажу ефикасан рад алата *Библиша* омогућавају да се на постављен упит добијени погоци прегледају и у виду конкорданци у текстовима часописа са тачно наведеним бројем, подацима о чланку и реченици/пасусу где се кључна реч из постављеног упита појављује. Конкорданце су, такође, двојезичне на српском и енглеском језику. Пример постављеног упита за појам који се већ налази у лексичким ресурсима дат је на слици 23, а део добијених конкорданци након постављеног упита види се на слици 24. Лексички ресурс „Библиотекарски термилошки речник” већ је садржао појам на српском и енглеском језику, упит је проширен само морфолошки, али не и семантички.

	Number of concordances (en): 89	Broj konkordansi (sr): 89
Brooks, 2009, vol. X:1/2, ID: 1.2009.1/2.3 metadata	n3 : Abstract: Through a detailed review of STM and business content available in Serbia, this article depicts the value of particular <i>databases</i> and the impact of a tremendous influx of information into the country in recent years.	n3 : Apstrakt: Kroz detaljan pregled elektronskog sadržaja iz oblasti prirodno-tehničkih nauka (STM) i biznisa dostupnog u Srbiji, ovaj članak opisuje vrednost pojedinih akademskih <i>baza podataka</i> i uticaj neverovatno velikog priliva podataka u zemlju, u poslednjih nekoliko godina.
Brooks, 2009, vol. X:1/2, ID: 1.2009.1/2.3 metadata	n12 : With the tremendous enhancements made in recent years, a new, closer look at these scholarly <i>databases</i> is not only essential, but will also prove valuable to anyone facilitating or conducting research.	n12 : Ako se imaju u vidu ogromna poboljšanja do kojih se došlo poslednjih godina, novo, pomnije razmatranje ovih naučnih <i>baza podataka</i> nije samo od suštinske važnosti, već će se pokazati kao veoma značajno za svakoga ko omogućava ili sprovodi istraživanja.
Brooks, 2009, vol. X:1/2, ID: 1.2009.1/2.3 metadata	n15 : A great deal of information availability (via <i>databases</i>) that has been realized by researchers in Serbia (and all around the world) may be attributed to a surprisingly large quantity and quality of full-text journals in both the sciences (STM) and the humanities & social sciences (including business).	n15 : Veliki deo dostupnosti informacijama (preko akademskih <i>baza podataka</i>), koju su istraživači u Srbiji (a i širom sveta) uvideli, može se pripisati iznenađujuće velikom kvantitetu i kvalitetu časopisa sa punim tekstovima. Ovo se tiče kako naučnih, tehničkih i medicinskih časopisa, tako i časopisa društvenih i humanističkih nauka (uključujući i biznis).
Brooks, 2009, vol. X:1/2, ID: 1.2009.1/2.3 metadata	n16 : Less than a decade ago, virtually no full-text physics journals were available via aggregated <i>databases</i> .	n16 : Pre manje od decenije, preko objedinjenih akademskih <i>baza podataka</i> , praktično da i nije bilo dostupnih časopisa sa punim tekstom iz oblasti fizike.
Brooks, 2009, vol. X:1/2, ID: 1.2009.1/2.3 metadata	n24 : In addition, the overall usage of EBSCO <i>databases</i> in Serbia is up nearly 10% from 2007 to 2008.	n24 : Osim toga, ukupno korišćenje EBSCO <i>baza podataka</i> u Srbiji je od 2007. do 2008. godine poraslo za 10%.
Brooks, 2009, vol. X:1/2, ID: 1.2009.1/2.3 metadata	n28 : The goal of full-text <i>databases</i> is not to replace existing subscriptions; it is to bring many new sources into the library at a reasonable cost.	n28 : Cilj <i>baza podataka</i> sa punim tekstom nije da zameni postojeće pretplate već da, po razumnoj ceni, uvede mnoge nove izvore.

Слика 24 – Пример добијених конкурданци у алату Библиша

Претрага

Захваљујући лексичким ресурсима и технологијама које су у њега уграђене, као и колекцији двојезичних текстова које база садржи, алат *Библиша* нуди претрагу према различитим критеријумима:

1. претрагу лексичких ресурса уносом кључне речи на српском или енглеском језику

Корисник уноси кључну реч у поље за претрагу и бира језик на ком је упит постављен (српски или енглески). Затим одабере колекцију текстова коју жели да претражује (све текстове или један или више часописа; у претраживањима у овој дисертацији користимо корпус текстова часописа „Инфотека”) и лексичке ресурсе који се користе приликом претраге (све или поједине). Претрага се даље врши кликом на дугме „Preview and modify terms for query.” Потом систем позива веб сервисе који проналазе синонине и термин на другом језику у лексичким ресурсима. Корисник може да обрише неке од понуђених термина ако их не сматра релевантним за своје претраживање, као и да допише оне које сматра релевантним, а који нису понуђени. У поступку дописивања користи се зарез „,” као сепаратор.³⁴¹ Проширени упит укључује

³⁴¹Ranka Stanković, Cvetana Krstev, Biljana Lazić and Dalibor Vorkapić. „A bilingual digital library for academic and entrepreneurial knowledge management.” In: J.C. Spender, Giovanni Schiuma and Vito Albino (eds.) *Proceeding of 10th International Forum on Knowledge Asset Dynamics — IFKAD 2015: Culture, Innovation and Entrepreneurship: connecting the knowledge dots*, Bari, Italy, 10-12 June 2015:1764-1777.

све синонине и преводе наведене на панелу при чему се направи дисјункција (OR) између термина на једном језику.

Ако као кључну реч унесемо појам „издаваштво” и одаберемо колекцију текстова часописа „Инфотека”, видећемо да wordnet на српском нуди синоним *издавачка делатност*, а на енглеском постоје у wordnet-у *publication, publishing*, а у Библиотекарском терминолошком речнику и *book publishing*. Коначан облик упита добија се морфолошким проширењем у ком се сваком изабраном термину уз помоћ технологија описаних у претходном одељку додају флективни облици – нпр. уз термин *издавачка делатност* придружена су морфолошка проширења *издавачке делатности, издавачкој делатности, издавачку делатност, издавачка делатности, издавачком делатности, издавачкој делатности*. Резултати претраге су исписани у виду линија корконданци у којима су сви пронађени термини на оба језика истакнути. На почетку сваке линије конкорданци стоји идентификација документа (чланка) из ког је конкретна корконданца. Уз идентификацију стоји и хипервеза „metadata” ка метаподацима на оба језика. Картица са метаподацима нуди хипервезе ка ТМХ колекцији и PDF документу са комплетним текстом.

У следећем примеру претраге уносом кључне речи, можемо уочити да је извршено семантичко и морфолошко проширење упита. Тражени појам на српском језику „препознавање именованих ентитета” није садржан ни у једном одабраном лексичком ресурсу (wordnet-у, Библиотекарском терминолошком речнику, *Библимиру*), као ни тражени појам на енглеском језику *named entity recognition* и његов акроним *NER*. Семантичким проширењем, приказаним на слици 25, односно уносом новог појма на српском и енглеском у поље *Библимир* и претрагом кроз комплетне текстове ТМХ колекције добили смо 28 корконданци у којима је за српски језик уочљиво морфолошко проширење у виду пронађених појмова *препознавању именованих ентитета, препознавања именованих ентитета, препознавањем именованих ентитета* (слика 26).

Слика 25 – Претрага семантички проширеним упитом у алату Библиша

Trtovac, 2013, vol. XIV:1, ID: 1.2013.1.8 metadata	n33 : Second day of the workshop started with presentation about NER by Lotte Wilms from the National Library of the Netherlands.	n33 : Drugog dana radionice je započeta prezentacijom Lote Vilms iz Nacionalne biblioteke Holandije o prepoznavanju imenovanih entiteta .
Trtovac, 2013, vol. XIV:1, ID: 1.2013.1.8 metadata	n34 : She explained that the project anticipates named entity recognition to persons, locations and organizations and showed examples of the mentioned named entities.	n34 : Ona je objasnila da je projektom predviđeno prepoznavanje imenovanih entiteta za osobe, lokacije i organizacije i dala primere za navedene imenovane entitete.
Trtovac, 2013, vol. XIV:1, ID: 1.2013.1.8 metadata	n37 : So far, this tool is available for named entity recognition in German, Dutch and French.	n37 : Za sada je ovaj alat dostupan za prepoznavanje imenovanih entiteta na nemačkom, holandskom i francuskom jeziku.
Trtovac, 2013, vol. XIV:1, ID: 1.2013.1.8 metadata	n38 : During the workshop, prof. Dr. Cvetana Krstev and prof. Dr. Dusko Vitas, representatives of the Group of Human Language Technologies, Faculty of Mathematics, University of Belgrade, reached agreement with the coordinator of the project, Hans-Jörg Lieder, to do named entity recognition for Serbian language of processed texts of Serbian historical newspapers too.	n38 : U toku radionice prof. dr Cvetana Krstev i prof. dr Duško Vitas, predstavnici Grupe za jezičke tehnologije Matematičkog fakulteta Univerziteta u Beogradu, dogovorili su sa koordinatorom projekta, Hansom-Jergom Liederom, da se na obrađenim tekstovima srpskih novina uradi i prepoznavanje imenovanih entiteta za srpski jezik.
Gucul-Milojević, 2010, vol. XI:1, ID: 1.2010.1.4 metadata	n45 : 1. Name Entity Recognition : recognition of the names of people and organizations, geographic names of places, time expressions and specific types of numeral expressions.	n45 : 1. Prepoznavanje imenskih entiteta: prepoznavanje imena ljudi i organizacija, naziva mesta, vremenskih izraza i određenih tipova numeričkih izraza.
Gucul-Milojević, 2010, vol. XI:1, ID: 1.2010.1.4 metadata	n253 : One of the most important problems in information extraction is named entity recognition .	n253 : Jedno od izuzetno važnih problema u ekstrakciji informacija je i prepoznavanje imenovanih entiteta .
Gucul-Milojević, 2010, vol. XI:1, ID: 1.2010.1.4 metadata	n254 : With the expansion of the area in which information extraction is applicable, the need for more precise recognition of named entities has been growing.	n254 : Širenjem oblasti u kojoj ekstrakcija informacija ima primenu raste potreba za preciznijim prepoznavanjem imenovanih entiteta .
Gucul-Milojević, 2010, vol. XI:1, ID: 1.2010.1.4 metadata	n255 : Three major topics have come into focus concerning the problem of named entity recognition : the recognition of proper names (personal names, names of organizations and locations), time expressions (date and time) and the expression of quantity (percentage and monetary values).	n255 : Problem prepoznavanja imenovanih entiteta fokusira se na tri velike oblasti: prepoznavanje vlastitih imena (ličnih imena, imena organizacija i lokacija), vremenskih izraza (datum i vreme) i izraza za količinu (procenata, novčanih vrednosti).

Слика 26 – Пример морфолошког проширења алатом Библиша

2. претрагу преко метаподатака на српском или енглеском језику

Претрага преко метаподатака налази се на страници „Metadata search“, а врши се у оквиру колекције на српском или енглеском језику, а могући критеријуми за

претрагу су име аутора, речи из наслова чланака, година издавања и кључна реч (критеријуми се користе појединачно или се комбинују Буловим операторима). Претрага преко метаподатака могућа је у оквиру једног или свих часописа. Ова врста претраге, као и пример претраге преко метаподатака биће детаљно представљени у наредном одељку.

3. прелиставање метаподатака

Прелиставање метаподатака на српском и енглеском језику за сваки појединачни часопис покреће се са странице „Metadata browse.” Са ове странице могуће је пронаћи жељени чланак и приступити ТМХ колекцији или документу са комплетним текстом у PDF формату.

Метаподаци

Да би претрага дигиталне библиотеке двојезичних часописа била што квалитетнија, алат *Библиша* омогућава прегледање и претраживање метаподатака о чланцима унутар електронског часописа, односно дигиталне библиотеке двојезичних (вишејезичних) часописа. Чланци су индексирани метаподацима тако да је претраживање могуће на српском и енглеском језику преко различитих параметара. Схему за метаподатке развила је Група за језичке технологије, као основа је послужило поједностављено TEI заглавље, а неке од етикета се користе по први пут (<creation id></creation id>). Атрибут <creation id> у елементу <teiheader> носи кодирану информацију о часопису и документу, а у елементу <tu> (translation unit) информацију о поглављу и одељку.

Метаподатке који се користе у алату *Библиша* можемо поделити у више целина.

1. Технички метаподаци– на првом месту су технички метаподаци, односно они који се односе на системско решење и њихов саставни део је и етикета <tei header></tei header>, као и поменута етикета (<creation id></creation id>):

```
<?xml version="1.0" encoding="UTF-8"?>
  <journals>
    <teiHeader>
      <fileDesc>
        <titleStmt>
          <title>Strategija za detekciju svojstava slozenih reci– imenice i pridevi</title>
        <respStmt>
          <resp>autor kolekcije</resp>
          <name>Cvetana Krstev, Aleksandra Trtovac</name>
        </respStmt>
        <respStmt>
          <resp>softversko rešenje</resp>
```

```

<name>Ranka Stanković</name>
</respStmt>
</titleStmt>
<publicationStmt>
  <p>Metapodaci o kolekciji paralelizovanih srpsko-engleskih dokumenata iz časopisa Infoteka. Paralelizacija se radi korišćenjem xAlign alata, generisanje txm datoteka korišćenjem LeXimir-a</p>
</publicationStmt>
<sourceDesc>
  <p><!-- Različite napomene, uglavnom vezane za implementaciju --Promenjen IDs za članke– u ID ušao i ID časopisa >>/p>
</sourceDesc>
<revisionDesc>Prva verzija april 2011. Časopisi: 2010.1, 2010.2</revisionDesc>
<revisionDesc>Druga verzija februar 2012. Dodati časopisi: 2009.1/2, 2011.1</revisionDesc>
<revisionDesc>Treća verzija 2014. Dodati časopisi Infoteka: 2011.2, 2012.1/2, 2013.1/2, 2014.1/2, Podzemni radovi: od 2004.13.1 do 2012.21.13, Arhitektura i urbanizam: od 2002.9.1 do 2013.37.1, Stomatološki glasnik Srbije 2012.1.1 i Tempus projekat BAEKTEL</revisionDesc>
</fileDesc>
</teiHeader>

```

2. Библиографски метаподаци о часопису– следећи сегмент метаподатака односи се на библиографске метаподатке о часопису (URL часописа и наслов часописа на српском и енглеском, као и ISSN часописа):

```

<JournalCollectionURL="www.unilib.bg.ac.rs/zajednica01/izdanja/infoteka" ID="1">
  <Aboutxml:lang="en" Desc="Journal of informatics and librarianship" Name="INFOtheca"/>
  <Aboutxml:lang="sr" Desc="Časopiszainformatikuibibliotekarstvo" Name="INFOteka"/>
  <ISSN>0354-6462</ISSN>
  ...

```

3. Структурални метаподаци и библиографски метаподаци о чланку– у наредном сегменту описују се поједини чланци из одређеног броја часописа. Они укључују структуралне (година, број, редни број чланка, категорија рада) и библиографске метаподатке (УДК, имена аутора, афилијација аутора, адреса њихове електронске поште, наслов чланка, сажетак, кључне речи, стране на којима се чланак налази, URL чланка), а сви метаподаци су на српском и енглеском језику:

```

<Document>
  <ID>2012.2.4</ID>
  <UDC>811.163.41'322.2 , 004.9:811.111'374</UDC>
  <Author>
    <ID>1</ID>
    <Name>Vesna Pajić</Name>
    <Mail>svesna@agrif.bg.ac.rs</Mail>
    <Institutionxml:lang="en">University of Belgrade, Faculty of Agriculture, Department for Agricultural Engineering</Institution>
    <Institutionxml:lang="sr">Univerzitet u Beogradu, Poljoprivredni fakultet, Institut za poljoprivrednu tehniku</Institution>
  </Author>
  <Author>

```

```

<ID>2</ID>
<Name>Staša Vujičić Stanković</Name>
<Mail>stasa@matf.bg.ac.rs</Mail>
<Institutionxml:lang="en">University of Belgrade, Faculty of Mathematics</Institution>
<Institutionxml:lang="sr">Univerzitet u Beogradu, Matematički fakultet</Institution>
</Author>
<Author>
  <ID>3</ID>
  <Name>Miloš Pajić</Name>
  <Mail>paja@agrif.bg.ac.rs</Mail>
  <Institutionxml:lang="en">University of Belgrade, Faculty of Agriculture, Department for
  Agricultural Engineering</Institution>
  <Institutionxml:lang="sr">Univerzitet u Beogradu, Poljoprivredni fakultet, Institut za
  poljoprivrednu tehniku</Institution>
</Author>
<Aboutxml:lang="en">
  <Title>Transducers for Annotating Weather Information in Meteorological Texts in
  Serbian</Title>
  <Category>Scientific paper</Category>
  <URLxml:lang="en">http://infoteka.bg.ac.rs/PDF/Eng/2012-
  2/INFOTHECA\_XIII\_2\_December2012\_33-47.pdf</URL>
</About>
<Aboutxml:lang="sr">
  <Title>Transduktori za označavanje podataka o vremenskim prilikama u meteorološkim
  tekstovima na srpskom jeziku</Title>
  <Category>Naučni rad</Category>
  <URLxml:lang="sr">http://infoteka.bg.ac.rs/PDF/Srp/2012-
  2/INFOTHECA\_XIII\_2\_December2012\_36-51.pdf</URL>
</About>
  <Abstractxml:lang="en">We present a process of extracting information on meteorological
  phenomena from texts in Serbian. We used finite state automata and transducers for both
  text processing and information extraction, through software specialized for linguistic text
  processing. Information extraction was done by annotating text segments...</Abstract>
  <Abstractxml:lang="sr">Uradu je prikazan jedan proces izdvajanja informacija o
  meteorološkim pojavama iz tekstova na srpskom jeziku. Obrada teksta, kao isamo izdvajanje
  informacija, vršeno je uz pomoć konačnih automata i transduktora, kreiranih i primenjenih
  pomoću programa specijalizovanih za lingvističku obradu teksta. Samo izdvajanje informacija
  vršeno je obeležavanjem segmenata teksta...</Abstract>
  <Keywordsxml:lang="en">information extraction, Serbian, natural language processing,
  finite state transducers, recursive transition networks</Keywords>
  <Keywordsxml:lang="sr">ekstrakcija informacija, srpski jezik, obrada prirodnih jezika,
  konačni transduktori, rekurzivne mreže prelaza</Keywords>
  <Pagesxml:lang="en">33a-47a</Pages>
  <Pagesxml:lang="sr">36-51</Pages>
</Document>
...
</journal>
</JournalCollection>
</journals>

```

Административни метаподаци – пошто часопис „Инфотека“ излази онлајн у режиму отвореног приступа, не постоје ограничења за коришћење, читање, цитирање и прегледање текстова, па тако не постоје ни рестрикције наглашене административним метаподацима.

У оквиру алата *Библиша* захваљујући метаподацима могућа је претрага преко имена аутора, речи из наслова, године издавања, кључних речи. После обављене претраге, подаци су видљиви у засебном прозору и у понуди је преглед преко целокупног текста чланка у PDF формату, као и ТМХ колекције документа са паралелним текстом на српском и енглеском језику „сравњеном“ до нивоа сегмента, односно реченице.

Слика 27 – Претрага преко метаподатака путем алата *Библиша*

Претрага преко метаподатака представљена је на слици 27, резултати претраге на слици 28, а преглед преко ТМХ колекције на слици 29.



Број докумената: 2, Језик: sr

Document	Authors	Title	Keywords
1.2008.1/2.7 tmx pdf	Aleksandra Vraneš	Iz istorije Katedre za bibliotekarstvo i informatiku Filološkog fakulteta Univerziteta u Beogradu	
1.2014.1.1 tmx pdf	Aleksandra Vraneš	Digitalna humanistika i savremene biblioteke	bibliotekarstvo, digitalizacija, digitalna humanistika, kultura, informacija

Слика 28 – Резултати претраге преко метаподатака

Недостаци коришћених лексичких ресурса

После извршене валидације алата *Библиша* утврђено је да претрага која укључује наведене ресурсе не даје увек добре резултате у проналажењу релевантних информација. Уколико термин који се претражује преко комплетног текста није унет ни у један од понуђених лексичких ресурса, неће се добити никакав одзив иако корпус текстова садржи тај термин.

English- (199)		Srpski - (199)	
n1 : Digital Humanities and Modern Libraries	n1 : Digitalna humanistika i savremene biblioteke	n2 : Sažetak:	n2 : Sažetak:
n3 : According to the Digital Humanities Manifesto 2.0, the first wave of digital humanities in the late nineties and early 21st century was primarily focused on large digitization projects and the establishment of technological infrastructure, which resulted in expecting the cooperation, interactivity, interdisciplinarity, multifunctionality, multimodality, dynamism, speed, accessibility and searchability from libraries as cultural and educational institutions.	n3 : Prema Manifestu digitalne humanistike (Digital Humanities Manifesto 2.0), prvi talas digitalnih humanističkih nauka krajem devedesetih godina prošlog veka i početkom 21. veka bio je prevashodno usredsređen na velike projekte digitalizacije i uspostavljanje tehnološke infrastrukture, zbog čega se od biblioteka kao institucija kulture i obrazovanja očekivala saradnja, interaktivnost, interdisciplinarnost, multifunkcionalnost, multimedijalnost, dinamičnost, brzina, dostupnost, pretraživost.	n4 : They are, along with museums, archives, theaters, cinemas, nowadays, as well as throughout the history, active agents of intercultural and interethnic communication.	n4 : One su, uz muzeje, arhive, pozorišta, bioskope, danas, kao i kroz istoriju, aktivni nosioci interkulture i međunacionalne komunikacije.
n5 : Primarily due to their general ethical and aesthetic values, characteristics of particular cultures that they are profiling, and then due to the standardized professional peculiarities that allow permeation of library science with other disciplines, as well as horizontal connectivity within the profession in the interest of the nation and its culture.	n5 : Pre svega, po opštim etičkim i estetskim vrednostima, po osobenostima zasebnih kultura koje profilisu, a potom po standardizovanim strukovnim posebnostima koje omogućavaju prožimanje bibliotekarstva sa drugim disciplinama, kao i horizontalno povezivanje u okviru struke u interesu naroda i njemu pripadajuće kulture.	n6 : Through forming a digital collection of our humanities, we are returning debt that we have towards its creators, previous generations, and commit the following generations to continue to preserve our cultural, spiritual and national identity.	n6 : Formirajući digitalnu zbirku naše humanistike, na opštem, lokalnom i institucionalnom nivou, vraćamo dug njenim tvorcima, prethodnim generacijama, i zadužujemo sledeće koje valja da nastave da čuvaju naš kulturni, duhovni i nacionalni identitet.
n7 : Keywords: librarianship, digitization, digital humanities, culture, information.	n7 : Ključne reči: bibliotekarstvo, digitalizacija, digitalna humanistika, kultura, informacija.	n8 : "History of libraries is the history of serving" this is how Frederick Friend, one of the initiators of The Scholarly Publishing and Academic Resources Coalition (SPARC) summed up his own and general knowledge.	n8 : "Istorija biblioteka je istorija služenja" - sazeo je svoje i opšte spoznaje Frederick Friend, jedan od inicijatora Koalicije univerzitetskih izdavača i otvorenih akademskih izvora (SPARC The Scholarly Publishing and Academic Resources Coalition).
n9 : "It is also the history of changes (...) there is only one value that is retained from the past in the new environment.	n9 : "Takođe, to je istorija promena (...) samo je jedna vrednost iz prošlosti zadržana i u novom okruženju.	n10 : The mission of the library can be simply explained as enabling services essential to users.	n10 : Misija biblioteke može se jednostavno objasniti kao omogućavanje usluga neophodnih korisnicima.
n11 : This mission is shared by all libraries in every generation and throughout the world."	n11 : Ovu misiju dele sve biblioteke u svakoj generaciji i u celom svetu."	n12 : Until recently, we approached the mission of the library in the traditional manner, and nowadays, for its fulfillment, we expect: collaboration, interactivity, interdisciplinarity, multifunctionality, multimodality, dynamism, speed, accessibility, searchability, hence, a set of these properties and activities led contemporary theorists to name such a library virtual or postmodern.	n12 : Do skora se misiji biblioteke pristupalo na tradicionalan način, a danas se za njeno ispunjavanje očekuju: saradnja, interaktivnost, interdisciplinarnost, multifunkcionalnost, multimedijalnost, dinamičnost, brzina, dostupnost, pretraživost, te skup ovih osobina i aktivnosti navodi savremene teoretičare da takvu biblioteku nazovu virtualnom ili postmodernom.

Слика 29 – Преглед паралелног текста чланка преко ТМХ колекције

На слици 30 види се пример такве претраге. Наиме, на упит „дигитална хуманистика” у оквиру колекције текстова часописа „Инфотека”, а коришћењем лексичких ресурса wordnet и „Библиотекарски термилошки речник” није добијен никакав одзив, односно термин није пронађен у лексичким ресурсима. Ако погледамо претрагу са претходне две слике (слике 28 и 29) уочићемо да се се термин „дигитална хуманистика” појављује бар у једном тексту у часопису „Инфотека.” Ипак, без обзира на то што се термин не налази у понуђеним лексичким ресурсима, тим термином можемо да претражујемо.

Управо због оваквих случајева приступило се креирању ресурса *Библимир* који треба да помогне у превазилажењу недостатака и унапреди перформансе алата *Библиша*.

Основни разлог за креирање новог ресурса лежи у чињеници да је од почетних коришћених ресурса само српски wordnet развила Група за језичке технологије Универзитета у Београду, а да су остали ресурси морали да се преузму такви какви јесу и нису могли да се мењају. Још један разлог је што се сматрало да ће се развојем новог ресурса алат *Библиша* боље радити јер ће лексички ресурси бити унапређени.

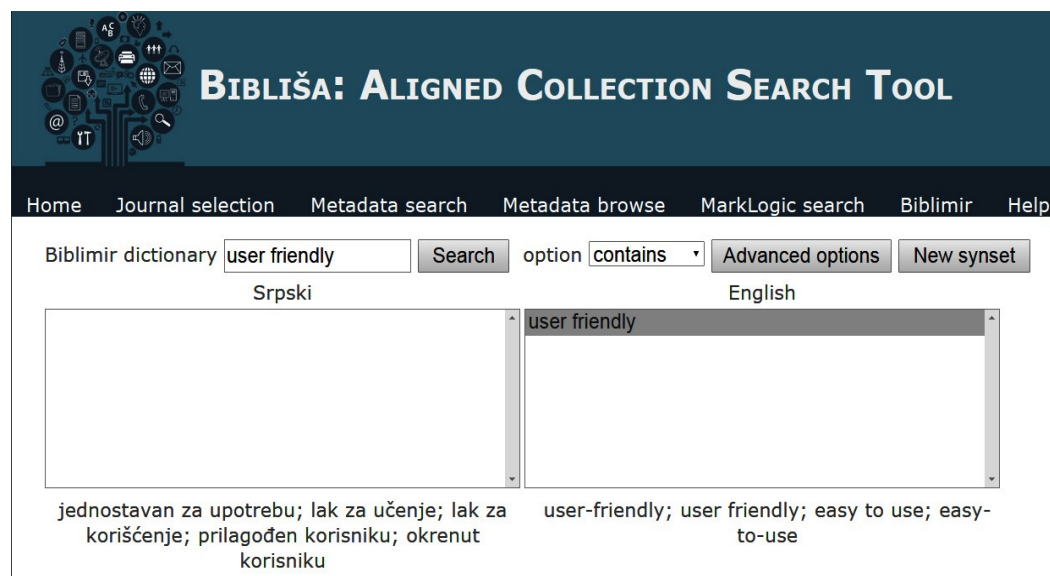
Слика 30 – Задавање претраге термина који се не налази у означеним лексичким ресурсима

Сврха ресурса *Библимир* огледа се на следећем примеру: у енглеском wordnet-у јавља се термин *user-friendly*, али у српском wordnet-у не постоји његов еквивалент. С друге стране, у „Библиотекарском терминолошком речнику“ постоји термин *user friendly* (без цртице) са преводима *једноставан за употребу*, *лак за учење и коришћење*. Пошто ови преводи термина нису у српском језику довољно прецизни, приликом уноса у ресурс *Библимир*, они су разложени на *једноставан за употребу*, *лак за учење*, *лак за коришћење*, *прилагођен кориснику*, *окренут кориснику*. Пошто за српски термин *лак за коришћење* у *Библимиру* постоје енглески преводи (са две ортографске варијанте) *easy to use*, *easy-to-use*, онда енглески синсет садржи следеће термине *user-friendly*, *user friendly*, *easy to use*, *easy-to-use*, а одговарајући српски синсет садржи *једноставан за употребу*, *лак за учење*, *лак за коришћење*, *прилагођен кориснику*, *окренут кориснику*.³⁴² На овај начин извршено је семантичко проширење упита, док се морфолошко проширење упита заснива на томе да систем констатује да одређене вишечлане речи нема у морфолошком речнику српског језика, али ће покушати да погоди како се та реч мења и то ће бити углавном тачно. Код погрешно

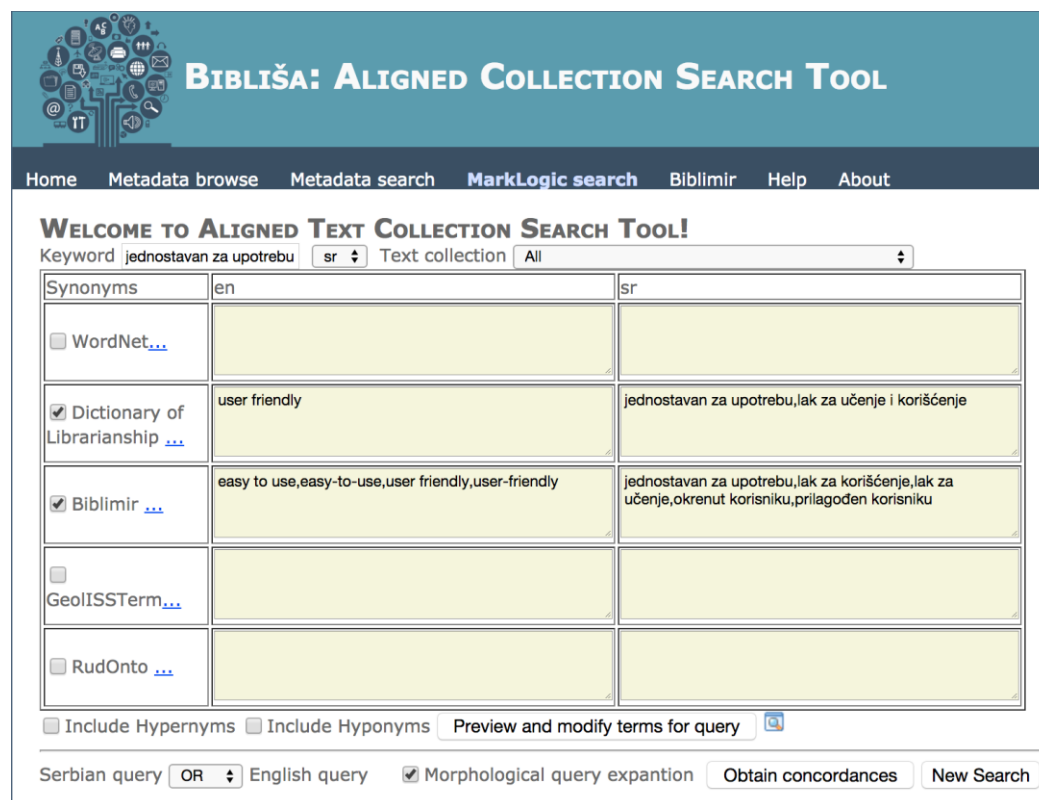
³⁴²Ranka Stanković, Ivan Obradović and Aleksandra Trtovac. „An Approach to Development of Bilingual Lexical Resources.” In: Zoran Budimac, Mirjana Ivanović and Miloš Radovanović (eds.) *Local Proceedings of the Fifth Balkan Conference in Informatics, 2012, BCI '12*, Novi Sad, Serbia, September 16-20, 2012:101-104, <http://ceur-ws.org/Vol-920> (приступљено 23.10.2014)

препознатих промена речи, врши се ручна корекција и исправан унос у морфолошки речник српског језика.

Пример претраге описаног термина приказан је на слици 31, а надоградња Библише видљива је на слици 32.



Слика 31 – Пример претраге термина коришћењем ресурса Библимир



Слика 32 – Пример претраге проширивањем упита алатом Библиша

Како је и описано, алат *Библиша*, поред недостатака, нуди бројне могућности у погледу претраге и екстракције информација из двојезичних (и вишејезичних) електронских часописа, али је и значајно средство за допуну морфолошких речника српског језика и српског и енглеског wordnet-а. У оквиру истраживања која су задата овом докторском дисертацијом, алат *Библиша* ће се користити у смислу доприноса развоју лексичких ресурса и проналажењу термина који нису саставни део морфолошких речника српског. Фокус у истраживању применом овог алата је на терминологији из области библиотекарства и информатике при чему ће се консултовати и поменути терминолошки речник. Нарочито је значајно утврдити који се термини не јављају ни у терминолошком речнику, ни у морфолошким речницима српског језика, али су садржани у паралелним текстовима чланака на српском и енглеском језику у часопису „Инфотека.”

5 Проналажење информација – анализа

У оквиру овог поглавља у наредним одељцима извршићемо анализу проналажења информација у библиотечким каталозима, универзитетским репозиторијумима и дигиталним библиотекама наведеним у одељку 1.4, а помоћу дескриптора метаподатака, као и дескриптора садржаја у случајевима где је таква претрага предвиђена и могућа. Како је један од циљева овог истраживања допуна терминолошког речника из библиотекарства и информатике вишечланим терминима, тако ће се и претрага кроз наведене корпусе заснивати на упитима који садрже вишечлане речи.

5.1 Примери претраге записа у каталозима применом дескриптора метаподатака

Библиотечки каталог Виртуелне библиотеке Србије (COBIB.SR) и каталог Европске библиотеке (TEL) садрже записе који су опремљени само дескрипторима метаподатака, али не и дескрипторима садржаја, па ће се претрага и проналажење информација на корпусу тих записа вршити преко метаподатака за записе на српском језику.

5.1.1 Претрага записа у каталогу COBIB.SR

Дескрипторима метаподатака у каталогу COBIB.SR предвиђена је претрага преко различитих критеријума који су у оквиру формата за унос библиографских података индексирани за претрагу – примарног, секундарног и алтернативног аутора, главног стварног наслова, упоредног наслова на другом језику, поднаслова тј. додатка наслову, године издавања, издавача, места издавања, године штампања, назива штампарије, места штампања, назива колекције, редног броја у колекцији, предметних одредница, врсте ауторства, УДК броја и сл.

Како је претрага вишечланим речима које спадају у домен терминологије из библиотекарства и информатике најчешће везана за метаподатке из наслова, предметних одредница и текстуалне описе УДК броја (ови подаци су у поступку индексирања обједињени заједничким критеријумом за претрагу – кључна реч), тако ћемо код постављања упита за проналажења ове врсте информација у командном претраживању користити префиксе за српски језик (LA=srp) и кључне речи (KW=), односно у изборном претраживању одабрати српски језик и кључну реч у менију за претрагу.

Пример 1 – универзитетска библиотека

Ако имамо намеру да преко кључне речи у каталогу COBIB.SR пронађемо информације о публикацијама које се тичу појма *универзитетска библиотека*, морамо поставити следећу стратегију – да бисмо добили све публикације у којима се у насловима појављује овај појам, морамо имати у виду да због флексије српског језика појам неће увек бити у номинативу, те морамо извршити скраћивања Клинијевом звездом не бисмо ли као одзив на претрагу добили тражени појам у било ком падежу. Овакав проблем немамо код метаподатака за предметне одреднице јер су оне увек у номинативу. Упит ћемо дефинисати на следећи начин:

LA=srp and KW=(univerzitetsk* and bibliotek*)

Командно претраживање даје 938 погодака (видети слику 33). Изборно претраживање преко избора српског језика и истих кључних речи из менија даје 3.469 погодака, што се може видети на слици 34.

Узајамна база података: COBIB.SR - Узајамна библиографско-каталожка база података COBIB.SR (Бр. записа: 2.932.106)

Резултати командног претраживања

Претражили сте: LA=srp and KW=(univerzitetsk* and bibliotek*) И Избор записа=Сва грађа [Ток претраживања](#)

Број пронађених записа: 938 [ПРОМЕНИ ЗАХТЕВ ЗА ПРЕТРАЖИВАЊЕ](#)

Сортирање по: COBIB.SR-ID [Записи: 1-10] 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 ... > >>

Бр.	Аутор	Наслов	Врста грађе	Језик	Година	Доступност фонда	Е-приступ
1.	Brzulović-Stanisavljević, Tatjana, 1965-	Дисеминација научних информација и институционални репозиторијум	излагање на конференцији	srp	2013	у публикацији	
2.	Бован, Владимир	Манојло Ђорђевић Призренац : живот и дело	књига	srp	2002	за позајмицу - у читаоници	
3.	Mošin, Vladimir, 1894-1987	Копитарева збирка и њен опис Владимира Мошина [Електронски извор]	рачунарски CD/DVD	srp	2007	за позајмицу - у читаоници	
4.	Orfelin, Zaharija, 1726-1785 Mišić, Smitilja	Поздрав Мојсеју Путнику [Електронски извор]: Нови Сад, 1757 : Рукопис у биб. у Вроцлаву	рачунарски CD/DVD	srp	2005	за позајмицу - у читаоници	
5.	Јанић, Ђорђе Ј.	Трагање за вером Исидоре Секулић	књига	srp	2001	за позајмицу - ван библиотеке	

Слика 33 – Резултати командног претраживања за тражену кључну реч *универзитетска библиотека* у каталогу COBIB.SR

Резултати изборног претраживања

Претражили сте: Кључне речи=универзитетск* И Кључне речи=библиотек* И Избор записа=Сва грађа И језик=srp [Ток претраживања](#)

Број пронађених записа: 3469 [ПРОМЕНИ ЗАХТЕВ ЗА ПРЕТРАЖИВАЊЕ](#)

Сортирање по: COBIB.SR-ID [Записи: 1-10] 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 ... > >>

Бр.	Аутор	Наслов	Врста грађе	Језик	Година	Доступност фонда	Е-приступ
1.	Роде, Борис	Mašinski materijali. Dio 1 i 3	учбеник	ssc	1983	за позајмицу - у читаоници	
2.	Вучковић, Жељко Иваз, Људмила	Народне и високошколске библиотеке у Војводини у 1992. години	чланак - саставни део	ssc	1993	у публикацији	
3.	Чаушевић, Мехмед	Statika konstrukcija po teoriji drugog reda. Dio 1	учбеник	ssc	1983	за позајмицу - ван библиотеке	
4.	Jaćimović, Branislav, 1951- Genić, Srbslav, 1963-	Топлотне операције и апарати	учбеник	ssc	1992	за позајмицу - ван библиотеке	
5.	Voronjec, Dimitrije Kuburović, Miloš	Проблеми из термодинамике вишекомпонентних система и хемијске термодинамике	учбеник	ssc	1991	за позајмицу - у читаоници	

Слика 34 – Резултати изборног претраживања за тражену кључну реч *универзитетска библиотека* у каталогу COBIB.SR

Запазићемо огромну разлику у броју погодака у истом каталогу, ако упит поставимо у различитим сегментима претраге.

Разлог за то јесте што је стари код за српски језик, према ISO стандардизацији био scc, а од 2012. године замењен је кодом srp. Међутим, аутоматска конверзија и замена старог кода новим кодом није извршена, па индекс за претрагу LA=srp проналази само оне публикације на српском језику за које је каталoшки запис креиран од 2012. године, без обзира на годину издавања публикације. Да бисмо претрагом обухватили све записе за публикације на српском језику, морамо у упит уврстити и стари код за српски језик. Нови упит у командном претраживању би изгледао овако:

LA=(srp or scc) and KW=(univerzitetsk* and bibliotek*)

На овако постављен упит добили смо одзив од 3.469 погодка у каталогу. Број погодака је сада идентичан броју погодака преко изборног претраживања. Међутим, уколико корисник није упућен да постоје два кода за српски језик у бази каталoшких записа, он неће извршити поуздану претрагу.

На слици 34 видимо листу добијених погодака. На првом месту је запис за публикацију која има наслов *Mašinski materijali*. Пошто се наш упит односио на универзитетске библиотеке, погледаћемо метаподатке за овај запис да бисмо установили зашто се он нашао међу погоцима (слика 35).

Узајамна база података: **COBIB.SR** - Узајамна библиографско-каталoшка база података COBIB.SR (Бр. записа: 2.932.106)

Изабрани запис [трајни линк](#)

Пуни **ISBD** **COMARC** запис [1/1000] > >>

Аутор	Роде, Борис
Наслов	Mašinski materijali. Deo 1 i 3 / Boris Rode ; [crteže izradio Bedri Džafa]
Врста/садржај	уџбеник
Језик	српски
Година	1983
Издавање и производња	Приштина : Zavod za udžbenike i nastavna sredstva Socijalističke Autonomne Pokrajine Kosova, 1983 (Bor : Bakar)
Остали аутори	Џафа, Бедри
Физички опис	286 str. : ilustr. ; 24 cm
Збирка	Biblioteka Univerzitetski udžbenici
Напомене	Tiraž 500 Bibliografija: str. 277-280.
ISBN	(Broš.)
Предметне одреднице	Машински материјали
УДК	621.77.9(075.8)
COBISS.SR-ID	36086535

Слика 35 – Испис комплетних метаподатака у каталогу COBIB.SR

Запазићемо да је описана публикација издата у оквиру збирке која се зове *Biblioteka "Univerzitetski udžbenici"*, те да је то једини разлог што је међу пронађеним записима информација о овој публикацији. На овом месту неопходно је објаснити да индекс за претрагу преко кључне речи подразумева и претрагу преко метаподатака за назив издавача, назив колекције, текста напомена и текста сажетка у случајевима кад је унет у формат за библиографске податке.

Ради анализе прецизности постављеног упита, узећемо у обзир само монографске публикације и ограничићемо временски период издавања публикација на оне које су издате од 2010. Овај упит даје 521 погодак, а на универзитетске библиотеке се односи само 41 публикација. Прецизност у проналажењу информација износи $P = \frac{41}{521} = 0,0787$ или 7,87%.

Закључићемо да претрага преко кључне речи не даје прецизне резултате на постављени упит, те да је неопходно променити упит за претраживање да бисмо добили информације о публикацијама које се односе на универзитетске библиотеке.

Промена упита и добијање прецизнијих погодака захтева додатно познавање могућности које нуди претрага каталога. Промењени захтев за претрагу изгледао би овако:

(LA=scc or srp) and SU=(univerzitetsk* and bibliotek*) and PY=201*

Префикс SU= у новом упиту односи се на све врсте предметних одредница. Није било могуће поставити критеријум да наслов садржи речи универзитетска и библиотека јер претрага преко наслова подразумева тачан унос наслова или његовог почетка. Када смо дефинисали да вршимо претрагу преко предметне одреднице, онда смо као одзив добили само оне публикације које су предметизоване на овај начин. Одзив је изузетно мали – 2 записа, али је прецизност потпуна – 100%. Обе публикације се односе на универзитетске библиотеке. Јавља се додатни проблем – субјективност предметизатора. Можда се неке публикације делимично односе на универзитетске библиотеке, али је предметизатор сматрао да је то недовољан део да би био истакнут метаподацима у предметној одредници. Како каталог COBIB.SR не нуди испис метаподатака са садржајем публикације, остајемо ускраћени за ову информацију.

Пример 2 – библиографски опис

У овом случају поставили смо упит LA=(srp or scc) and kw=(bibliografsk* and opis*) и ограничили претрагу на монографске публикације. Дobili смо 54 погодка као одзив на постављени упит. Резултати претраживања приказани су на слици 36.

Резултати командног претраживања

Претражили сте: LA=(srp or scc) and kw=(bibliografsk* and opis*) [Избор записа=књиге](#) [Ток претраживања](#)

Број пронађених записа: 54 ПРОМЕНИ ЗАХТЕВ ЗА ПРЕТРАЖИВАЊЕ

Сортирање по: COBISS.SR-ID [Записи: 1-10] 1 | 2 | 3 | 4 | 5 | 6 > >>

Бр.	Аутор	Наслов	Врста грађе	Језик	Година	Доступност фонда	Е-приступ
1.	Цисарж, Бранко А.	Један век периодичне штампе Српске православне цркве : (библиографски опис часописа и листова са прегледом-садржајем свих радова објављених у њима од 1868-1970). [Књ. 1, А-М]	књига	srp	1986	није за позајмицу	
2.	Цисарж, Бранко А.	Један век периодичне штампе Српске православне цркве : (библиографски опис часописа и листова са прегледом-садржајем свих радова објављених у њима од 1868-1970). [Књ. 1, А-М]	књига	srp	1986	није за позајмицу	
3.	Cisarž, Branko A.	Један век периодичне штампе Српске православне цркве : библиографски опис часописа и листова са прегледом-садржајем свих радова објављених у њима 1868-1970. 2. Н-Ш	књига	scc	1986	за позајмицу - ван библиотеке	

Слика 36 – Резултати командног претраживања за тражену кључну реч *библиографски опис* у каталогу COBISS.SR

Увидом у релевантност погодака, утврдили смо да се 29 публикација односи на задати појам у упиту, те да је прецизност $P = \frac{29}{54} = 0,54$ или 54,7%. Резултати су у овом проналажењу информација неупоредиво бољи него у претходном, али и даље само половина добијених резултата одговара захтеву за претрагу.

Уколико уместо кључних речи, захтев за претрагу дефинишемо преко предметних одредница, добићемо следећи резултат – само три записа у метаподацима садрже предметну одредницу *библиографски опис* иако се, без сумње, 29 публикација односи на ову тему што значи да је одзив изузетно мали. Предметизација ових публикација није узела у обзир тај узак појам већ су се предметизатори определили за предметне одреднице *библиотечки стандарди*, *каталогизација*, *библиотечки фондови* и сл.

Пример 3 – архивска грађа

Код следеће претраге намера је била да пронађемо публикације на српском језику које говоре о архивској грађи. Поново смо ограничили претрагу само на монографске публикације и поставили следећи упит:

LA=(srp or scc) and KW=(arhivsk* (W) građ*).

На постављени упит добили смо одзив од 570 погодака. Већ код првог погодка, установили смо да публикација не говори о архивској грађи јер је наслов публикације „Ljudi i lože: ruski slobodni zidari XX veka.” Детаљним увидом у пуне метаподатке о запису, уочили смо да се појам *архивска грађа* јавља у тексту једне од напомена. Како индекс KW= врши претрагу и преко текста напомена, тако је запис за ову публикацију одабран као погодак.

На листи погодака на 21. месту јавио се запис за публикацију „Ариљска кинологија.” Сам наслов публикације говори да она вероватно није повезана са темом

архивска грађа. Прегледањем пуног записа утврдили смо да публикација садржи и архивску грађу о Кинолошком друштву „Ариље“, па је то разлог зашто се ова публикација нашла међу погоцима као одзив на постављени упит.

На основу примера ова два погодка, јасно је да је било неопходно променити критеријуме за претрагу. Нова претрага је вршена преко следећег упита:

LA=(srp or scc) and SU=arhivska građa.

Узајамна база података: COBIB.SR - Узајамна библиографско-каталожна база података COBIB.SR (Бр. записа: 2.932.106)

Резултати командног претраживања

Претражили сте: LA=(srp or scc) and SU=arhivska građa И Избор записа=књиге [Ток претраживања](#)

Број пронађених записа: 279 [ПРОМЕНИ ЗАХТЕВ ЗА ПРЕТРАЖИВАЊЕ](#)

Сортирање по: COBISS.SR-ID [Записи: 1-10] 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 ... > >>

Бр.	Аутор	Наслов	Врста грађе	Језик	Година	Доступност фонда	Е-приступ
1.		Списи бечких архива о Првом српском устанку. књ. 11, 1811	књига	scc; lat; ger	1971	нема фонда	
2.		Научни skup Борбе за ослобођење Крајева и Саска 1944. godine	књига	scc	1986	● за позајмицу - у читаоници	
3.	Архив Југославије Београд	Lista kategorija registraturskog materijala sa rokovima čuvanja Narodne banke Jugoslavije	књига	scc	1991	● за позајмицу - ван библиотеке	
4.	Југославија	Zakon o arhivskoj građi Federacije	закони	scc	1986	● за позајмицу - ван библиотеке	
5.	Југославија	Opšti zakon o arhivskoj građi	закони	scc	1964	● за позајмицу - ван библиотеке	

Слика 37 – Резултати командног претраживања за тражену предметну одредницу *архивска грађа* у каталогу COBIB.SR

Претрагу смо поново ограничили на монографске публикације, а уместо кључних речи користили смо индекс за претрагу свих предметних одредница. Добили смо одзив од 279 погодака (видети слику 37). Анализирајући листу погодака, утврдили смо да је код само 66 записа тема публикације архивска грађа, а да је у осталим публикацијама садржана архивска грађа о некој установи, појединцу, држави догађају и сл. Прецизност у проналажењу информација у овом случају је $P = \frac{66}{279} = 0,24$ или 23,7%. Закључили смо да је нужно да се претрага изврши искључиво преко тематске предметне одреднице (префикс TN=), уз ограничавање претраге на монографске публикације на српском језику.

Поставили смо нови упит:

LA=(srp or scc) and TN=arhivska građa.

На овако постављен упит добили смо тачно 66 погодака. Провером смо утврдили да се сваки од записа за публикације односи на тему „архивска грађа“, па је прецизност пронађених информација 1 или 100%.

На основу ова три примера можемо закључити да је за претрагу узајамног каталога Виртуелне библиотеке Србије COBIB.SR потребна велика вештина у коришћењу префикса за претраживање и формирању упита, као и да је и поред тога могуће

добити доста нерелевантних докумената као одзив на постављен упит. Претраживање се може извести и на једноставан начин, уносом кључних речи, али ће се најчешће добити велики број нерелевантних информација које је потом потребно искључити из листе потребних докумената.

Такође, непостојање контролисаних речника, тезауруса и усвојених предметних одредница доводи до субјективности предметизатора који врши садржинску обраду публикације, па две особе могу различито предметизовати исту публикацију што се види из примера записа за једну исту публикацију у локалним каталозима Народне библиотеке Србије (НБС) и Универзитетске библиотеке „Светозар Марковић“ (УБСМ) приказаним на слици 38.

Локална база података: Народна библиотека Србије, Београд (Бр. записа: 1.018.934) 🏠			
Изабрани запис трајни линк			
Пуни	ISBD	COMARC	запис [1/1]
Наслов	Smjernice za projekte digitalizacije zbirki i fondova u javnom vlasništvu, posebno onih koji se čuvaju u bibliotekama i arhivima / [prevodilac Vesna Vučković]		
Јединствени наслов	Guidelins for Digitization Projects for Collections and Holdings in the PublicDomain, Particularly Those held by Libraries and Archives. scc		
Врста/садржај	приручник		
Језик	српски		
Година	2012		
Издавање и производња	Cetinje : Nacionalna biblioteka Crne Gore "Đurđe Crnojević", 2012		
Остали аутори	Вучковић, Весна		
Физички опис	98 str. ; 24 cm		
Збирка	Priručnici / [Nacionalna biblioteka Crne Gore "Đurđe Crnojević"] ; knj. 1		
Напомене	Prevod djela: Guidelines for Digitization Project for Collections and Holdings in the Public Domain, Particularly those Held by Libraries and Archives "Ove Smjernice ja sačinila radna grupa sastavljena od predstavnika IFLA-e i ICA-e kojoj je UNESKO povjerio zadatak izrade smjernica za programe digitalizacije zbirki i građe u javnom vlasništvu." --> Predgovor. Napomene i bibliografske reference uz tekst. Bibliografija uz svako poglavlje.		
ISBN	ISBN 978-86-7079-129-9 (broš.)		
Предметне одреднице	Библиотека грађа - Дигитализација Архивска грађа - Дигитализација		
УДК	025.85:004.056.5 930.253:004.056.5(082)		
COBISS.SR-ID	21232400		

Локална база података: Универзитетска библиотека 'С. Марковић', Београд (Бр. записа: 302.642) 🏠			
Изабрани запис трајни линк			
Пуни	ISBD	COMARC	запис [1/1]
Наслов	Smjernice za projekte digitalizacije zbirki i fondova u javnom vlasništvu, posebno onih koji se čuvaju u bibliotekama i arhivima / [Prevodilac Vesna Vučković ; urednik Jelena Đurović]		
Јединствени наслов	Guidelins for Digitization Projects for Collections and Holdings in the PublicDomain, Particularly Those held by Libraries and Archives. scc		
Врста/садржај	приручник		
Језик	српски		
Година	2012		
Издавање и производња	Cetinje : Nacionalna biblioteka Crne Gore "Đurđe Crnojević", 2012 (Podgorica : Pobjeda)		
Остали аутори	Vučković, Vesna		
Физички опис	98 str. ; 24 cm		
Збирка	Priručnici ; knj. 1		
Напомене	Prevod djela: Guidelines for Digitization Project for Collections and Holdings in the Public Domain, Particularly those Held by Libraries and Archives Tiraž 500.		
ISBN	ISBN 978-86-7079-129-9 (broš.)		
Предметне одреднице	Digitalizacija - Bibliotečka građa - Priručnici Digitalizacija - Arhivska građa - Priručnici		
УДК	005.922.52:930.25(035) 005.922.52:027.85		
COBISS.SR-ID	21232400		

Слика 38 – Пример различите садржинске обраде исте публикације у локалним каталозима НБС и УБСМ

Морамо нагласити да је ситуација сасвим другачија ако се претрага врши преко наслова публикације и имена аутора. Ако као критеријуме за претрагу унесемо следеће податке AU=Vraneš, Aleksandra* and TI=Visokoškolske biblioteke, добићемо 2 погодка. Обе публикације јесу оне које смо желели да пронађемо, само је реч о два различита издања – 1. изд. из 2004. и 2. изд. из 2010. године (видети слику 39).

Иза имена аутора смо приликом постављања упита ставили Клинијеву звезду. Разлог за то је различито вођење одреднице у узајамном каталогу. Почев од 2012, година рођења аутора је обавезан податак који се, приликом каталогизације публикације, уноси у оквиру метаподатака о ауторству.

Узајамна база података: **COBIB.SR** - Узајамна библиографско-каталожка база података COBIB.SR (Бр. записа: 2.932.106)

Резултати командног претраживања

Претражили сте: AU=Vraneš, Aleksandra* and TI=Visokoškolske biblioteke И Избор записа=Сва грађа [Ток претраживања](#)

Број пронађених записа: 2 [ПРОМЕНИ ЗАХТЕВ ЗА ПРЕТРАЖИВАЊЕ](#)

Сортирање по: COBIB.SR-ID [Записи: 1-2]

Бр.	Аутор	Наслов	Врста грађе	Језик	Година	Доступност фонда	Е-приступ
1.	Вранеш, Александра, 1960-	Високошколске библиотеке	научна монографија	srb	2010	дезидерат	
2.	Vraneš, Aleksandra, 1960-	Високошколске библиотеке	научна монографија	scs	2004	за позајмицу - ван библиотеке	

Слика 39 – Приказ резултата командног претраживања у каталогу COBIB.SR преко имена аутора и наслова публикације

Разлог за то је израда нормативне датотеке аутора. Међутим како нисмо сигурни да ли за конкретног аутора постоји запис у нормативној датотеци, а у понуђеној листи аутора са истим именом постоји облик и без године рођења (погледати пример на слици 40), Клинијева звезда је неопходан елемент приликом постављања упита. То је још нешто што корисник треба да зна да би успешно претраживао поменути каталог.

Узајамна база података: **COBIB.SR** - Узајамна библиографско-каталожка база података COBIB.SR (Бр. записа: 2.932.106)

Преглед појмова

по: "AX" од: "vraneš, aleksandra"

НАПОМЕНА: Клик на појам у табели садржај се преноси у поље за претраживање

Фреквентност	Аутор
166	VRANEŠ, ALEKSANDRA
1	VRANEŠ, ALEKSANDRA B.
2	VRANEŠ, ALEKSANDRA B., 1960-

Слика 40 – Различити облици имена истог аутора у каталогу COBIB.SR

Иако је прецизност приликом ове врсте претраге најчешће 1 или 100% (као и у анализираном примеру), недостатак је и тај што је потребно знати макар приближно тачан наслов публикације, а апсолутно је неопходно знати тачан почетак наслова.

5.1.2 Претрага записа у каталогу TEL – Европска библиотека

Каталог Европске библиотеке садржи записе из националних каталога европских библиотека, као и записе за дигиталне колекције које су настале и развијале се у

оквиру пројеката Европеана. У овом делу претраге, бавићемо се проналажењем информација у записима из обједињених националних каталога европских библиотека.

Аналогно претходној претрази и резултатима представљеним у одељку 5.1.1, и у овом случају ћемо анализирати претрагу истих појмова.

Пример 1 – универзитетска библиотека

И на порталу Европске библиотеке поставили смо упит којим смо желели да пронађемо публикације о универзитетским библиотекама. На основном претраживању на упит *univerzitetsk*+bibliotek** добили смо 2.134 погодака. Желели смо да нашу претрагу ограничимо само на записе на српском језику па смо у менију језика одабрали српски и добили 189 погодака (видети слику 41). Приступили смо евалуацији понуђених записа за публикације и утврдили да се 76 публикација не односе на тему *универзитетска библиотека*, већ издавач у свом називу има речи *универзитетска* и *библиотека*. До проналажења ових записа је дошло због тога што основни ниво претраге проналази и метаподатке у оквиру етикете <publication>. На основу ових података, прецизност у проналажењу информација је $P = \frac{76}{189} = 0,40$ или 40,2%.

The screenshot shows the search interface of The European Library. At the top, there is a search bar with the query *univerzitetsk*+bibliotek** and a 'GO' button. Below the search bar, there are navigation tabs: 'HOME', 'DISCOVER', and 'ACCESS OUR DATA'. The search results are displayed in a grid format. On the left, there are filters for 'Everything', 'Newspapers', 'Collections', 'Full text', and 'Remote Search'. Below these filters, there are 'Related Concepts' for 'Narodna i univerzitetska biblioteka Bosne i Hercegovine (Organization)', 'Biblioteka Popullore dhe Universitare e Kosovës (Organization)', and 'Народна и универзитетска библиотека "Климент Охридски" (Скопје) (Organization)'. A message indicates that 20 items out of 189 are shown for the query in Serbian. A result is visible: 'Joca Vujić, veliki dobrotvor Beogradskog univerziteta i Univerzitetske biblioteke' by Danica Filipović, published in 2012. On the right, there is a 'REFINE' section with a search bar and a 'GO' button, and a list of filters: 'Language: Serbian', 'Accessibility' (At Contributor: 169, Access Online: 22, Thumbnail Available: 14), 'Discipline', 'Language', 'Contributor', and 'Collection'.

Слика 41 – Резултати претраге записа на српском језику на порталу TEL-а на упит *универзитетск* библиотека**

Треба напоменути да је претрага резултирала у неким случајевима и са два и више записа за исту публикацију. Разлог за то је што систем Европске библиотеке не претпоставља препознаје дупликате. Ако дупликате урачунамо као нерелевантне




погодке, онда прецизност претраге опада. Како је у међу понуђеним погоцима 18 дупликата, онда је само 58 релевантних публикација, па је прецизност $P = \frac{58}{189} = 0,30$.




У следећем кораку смо користили напредно претраживање и извршили претрагу преко следећих параметара: предмет публикације је универзитетска библиотека и језик публикације је српски. Добили смо 26 погодака. Примећујемо да је број погодака изузетно низак. До оваквог одзива дошло је због тога што националне европске библиотеке предметизацију врше на језику свог каталогског центра, па се српски језик не јавља међу овим подацима, осим у случају библиотека које су метаподатке о предмету публикације пренеле из домаћег узајамног каталога. Да бисмо добили све публикације којима је тема универзитетска библиотека, а издате су на српском језику, било би неопходно да претрагу преко метаподатака о предмету публикације вршимо на различитим језицима, што је заправо немогуће.


VIAF
Virtual International Authority File


Search



Select Field: Select Index: All VIAF Search Terms: **Search**


Univerzitetska biblioteka "Svetozar Marković"   

Univerzitetska biblioteka "Svetozar Marković" (Beograd)   

Univerzitetska biblioteka Svetozar Marković Belgrade 

Univerzitetska biblioteka "Svetozar Marković" (Belgrad). 

Универзитетска библиотека Светозар Марковић  

Univerzitieska biblioteka Svetozar Marković (Belgrade) 

VIAF ID: 122872393 (Corporate)
Permalink: <http://viaf.org/viaf/122872393>
ISNI: [0000 0001 2097 1531](https://www.isni.org/0000-0001-2097-1531)

Слика 42 – Нормативни запис за Универзитетску библиотеку „Светозар Марковић” у бази VIAF

Позитивна карактеристика претраге на порталу Европске библиотеке је та што се након извршеног претраживања појављује опција која указује на садржаје који су сродни оним из постављеног упита – related concepts. На слици 41 видимо део понуђених сродних садржаја за наш упит. Прегледом целокупне понуђене листе садржаја који се односе на постављени упит утврдили смо да је међу њима и Универзитетска библиотека „Светозар Марковић.” Изабрали смо ту везу и добили списак докумената који се односе на поменуту установу. Осим ове везе, могуће је погледати запис за исту установу у нормативној бази VIAF која је саставни део Повезаних отворених података, те утврдити у којим се све облицима јавља назив

наведене установе. Из записа у бази VIAF омогућен је приступ до базе ISNI³⁴³ – Међународни стандардни идентификатор имена што је приказано на слици 42.

Утврдили смо и да метаподаци у оквиру записа у бази VIAF под опцијом „спољашње везе”, односно ”external link” нуде додатне информације о Универзитетској библиотеци „Светозар Марковић”, овог пута у виду чланака на Википедији. Један од чланака је и чланак на српском језику. Такође, постоји и веза ка запису у бази WorldCat Identities где је поменута библиотека анализирана из угла уређивања и публикавања сопствених издања.³⁴⁴ Могућност да се отварањем веза у записима добијају нови метаподаци о одређеном документу везана је за карактеристике семантичког веба и намеру креатора Европеане и Европске библиотеке да се подаци о једном документу повежу са свим сродним метаподацима. У ситуацији када претрага резултира добрим одзивом и високом прецизношћу добијених погодака, ова особина је изузетно корисна. Међутим, код недовољно високог одзива и, што је још битније, код ниске прецизности, овакве напредне могућности не могу доћи до изражаја.

Пример 2 – библиографски опис

У следећој претрази у оквиру основног претраживања поставили смо следећи упит: bibliografsk* opis*. Добили смо 672 погодака. Претрагу смо потом сузили и одабрали српски језик, те добили само 14 погодака (видети слику 43). Пошто је назив за овај термин идентичан на различитим језицима, први корак претраге укључио је и записе на словеначком, хрватском и македонском језику. Међу погоцима се појављују и записи на руском, енглеском, француском, италијанском, немачком и другим језицима. Разлог за то је што метаподаци о предметним одредницама у етикети <subject> садрже тражени појам на српском, хрватском или словеначком језику. На овом месту, могли бисмо рећи да је прецизност $P = \frac{14}{672} = 0,02$ или 2,1%

³⁴³ ISNI – International Standard Name Identifier, доступно на: <http://www.isni.org/>

³⁴⁴ WorldCat Identities. Univerzitetska biblioteka “Svetozar Marković”, доступно на: <http://www.worldcat.org/identities/lccn-n80-14562/>

Слика 43 – Резултати претраге записа на српском језику на порталу TEL-а на упит *библиографск опис***

Међутим, приступили смо даљој анализи и утврдили да се записи за 4 понуђене публикације уопште не односе на тражени појам, већ се појам јавља међу метаподацима у етикетама <description> или <publication>. Такође, поново се јављају дупликати записа – за један наслов постоји четири записа, а за други два записа. Када смо одбацили дупликате, утврдили смо да смо претрагом пронашли информације о само 6 различитих публикација на српском језику чија је тема библиографски опис. Прецизност је, заправо $P = \frac{6}{672} = 0,009$ или <1%.

Покушали смо да добијемо боље резултате и повећамо прецизност променом критеријума за претрагу. Нову претрагу смо извршили у командном претраживању и претрагу извршили преко критеријума за предмет публикације (subject) уз ограничавање резултата претраге на оне публикације које су на српском језику. Нисмо добили ниједан погодак. У претходној претрази смо утврдили да постоји 6 публикација на српском језику којима је тема библиографски опис.

Потребно је да објаснимо зашто је одзив у овом проналажењу информација једнак нули. Метаподаци у оквиру етикете <subject> садрже податке из сегмента предметних одредница унетих у каталожке записе у бази COBIB.SR. У већини записа за публикације о библиографском опису, предметна одредница је *каталогизација*, а не *библиографски опис*, па је то разлог за непостојање одзива на постављени упит.

Када смо искључили критеријум српски језик, добили смо 27 погодака за публикације на различитим језицима. Увидом у листу погодака, утврдили смо да су у питању записи за публикације које је Европској библиотеци доставила Национална и свеучилишна књижница у Загребу. Како су записи за те публикације у метаподацима у етикети <subject> садржале тражени појам, тако се ови записи јављају као одзив на постављени упит са прецизношћу 1.

Пример 3 – архивска грађа

Основним претраживањем на постављени упит *arhivsk* grad** добили смо 2.411 погодак. Након сужавања претраживања само на записе за публикације на српском језику добили 53 погодка (слика 44). Анализом резултата проналажења информација утврдили смо да је тема само 10 публикација архивска грађа, односно да постоји само 10 релевантних докумената, а да остале публикације садрже опис архивске грађе о институцијама, знаменитим личностима или државним пословима, односно да су нерелевантни документи за ово претраживање. Прецизност износи само $P = \frac{10}{53} = 0,19$ или 18,9%.

The screenshot shows the TEL portal search results for the query "arhivsk* grad*" in Serbian. The interface includes a search bar with the query, a "GO" button, and a "REFINE" sidebar. The search results are displayed in a list format, showing the title, source, and contributor for each item. The first result is "Arhivska građa o Narodnoj biblioteci u Beogradu : 1821- 1944" from the National Library of France. The second result is "Biblioteka i arhivska građa Profesorskog društva u Oficirskom zarobljeničkom logoru Oflag VI C u Osnabriku" from The British Library.

Слика 44 – Резултати претраге записа на српском језику на порталу TEL-а на упит *архивск* грађ**

Очекивали смо да ћемо напредном претрагом преко метаподатака о предмету публикације добити боље и прецизније резултате, али је одзив на постављени упит био 0. Како смо претрагом у COBIB.SR на исти упит добили 66 погодака са прецизношћу 1, очекивали смо сличан резултат и у каталогу TEL. Претрагом на порталу Европске библиотеке и проверама функционалности веза ка каталогу Народне библиотеке Србије, утврдили смо да сервер тренутно није у функцији,³⁴⁵ те да су из претраживања тако искључени сви потенцијални записи из домаћег каталога.

Закључићемо да је претрага у каталогу Европске библиотеке резултирала често високим одзивом, али да је прецизност ниска чак и када укључимо напредну претрагу. Неспорно је да је овај каталог огромна и драгоцен колекција која нуди претрагу метаподатака о најразличитијим документима са могућношћу ограничавања године, језика, научне области, збирке, установе и др. Ипак, пре било каквог претраживања треба добро осмислити стратегију претраге, али и проверити базичне техничке услове

³⁴⁵ Стање током фебруара 2016. Из Народне библиотеке Србије потврђено је да је у питању тренутни технички проблем.

– везу са сервером. Додатно, велики проблем је и неуједначеност у додели метаподатака између различитих учесника у систему, као и непостојање тезауруса за избор метаподатака који би елиминисали ту недоследност.

5.2 Примери претраге записа у репозиторијумима применом дескриптора метаподатака и дескриптора садржаја

У овом одељку представићемо резултате проналажења информација у оквиру два институционална репозиторијума – Дигиталног репозиторијума Универзитета у Београду (PHAIDRA) и Дигиталне библиотеке Филолошког факултета. Записи у ова два репозиторијума опремљени су како дескрипторима метаподатака, тако и дескрипторима садржаја. У примерима претраге, најпре ћемо анализирати одзив и прецизност пронађених докумената преко дескриптора метаподатака, а потом и преко дескриптора садржаја постављајући идентичне упите за претраживање.

5.2.1 Претрага записа у репозиторијуму PHAIDRA

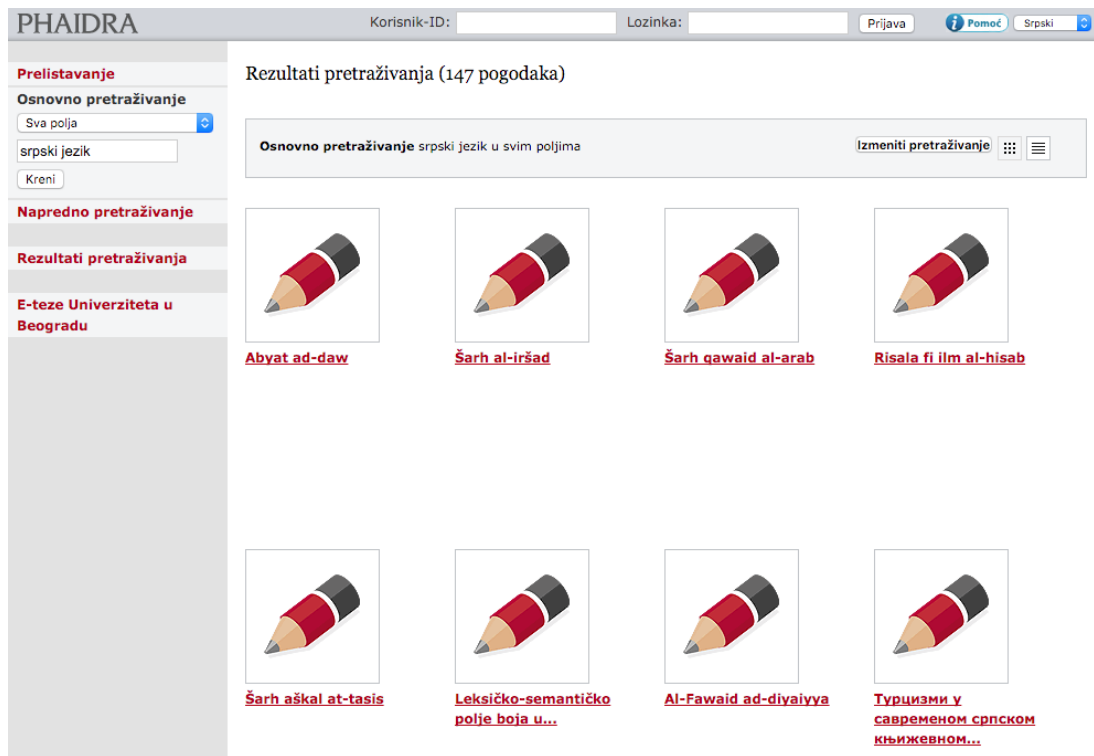
Репозиторијум PHAIDRA, као и већина база података, омогућава претрагу преко основног и преко напредног претраживања. Основно претраживање има сужен избор критеријума за претрагу (сва поља, идентификатор, наслов, опис, аутор и наслов књиге) док напредно претраживање нуди комбиновање различитих критеријума коришћењем Булове алгебре и укључивање (оператори AND и OR), али не и искључивање (оператор NOT) одређених критеријума.

Пример 1 – српски језик

У основно претраживање унели смо критеријум за претрагу *српски језик* и одабрали претрагу у свим пољима. Добили смо 147 погодака. На првој страни са добијеним резултатима, приказаној на слици 45, постоји 8 погодака. Само 2 од њих се односе на српски језик – реч је о две докторске дисертације чија је тема српски језик у различитим контекстима. Осталих 6 записа тичу се збирке оријенталних рукописа из фонда Универзитетске библиотеке „Светозар Марковић.”

Увидом у метаподатке о тих 6 рукописа, утврдили смо да су се ови записи нашли међу погоцима зато што је индексирање претраге по свим пољима одабрало и оне записе које у етикети <dc:description> садрже опис на српском језику. Већ на основу прве стране са погоцима, јасно је да се критеријум за претраживање мора променити. Ако из менија уместо опције *сва поља* одаберемо опцију *опис* резултати претраге су лошији. Свих 8 погодака на првој страни односе се на оријенталне рукописе.

Закључили смо да основна претрага за постављени упит не може дати релевантне погодке, па смо прибегли напредном претраживању.



Слика 45 – Резултати претраге записа у репозиторијуму PHAIDRA на упит *српски језик*

У напредној претрази смо дефинисали да желимо да нађемо документе који садрже кључну реч *српски језик* и који су у формату PDF, а документ је на српском језику. Добили смо 24 погодка. Првих осам погодака и последња четири у потпуности се односе на постављени упит. Међутим, осталих 12 погодака поново је везано за оријенталне рукописе. Разлог за то је што индекс за кључну реч поново повлачи метаподатке на српском језику из етикете <dc:description>. Такође, претрага је потпуно игнорисала захтев да документи буду на српском језику. Прецизност је у овом случају $P = \frac{12}{24} = 0,5$ или 50%.

Из свих ових разлога променили смо критеријуме за претрагу тако што смо додали критеријум да документ у поднаслову садржи фразу *докторска дисертација*. За овај критеријум смо се определили знајући да су докторске дисертације одбрањене на Универзитету у Београду после 2012. саставни део овог репозиторијума, па смо очекивали велику прецизност у пронађеним документима. Добили смо 8 погодака. Сви погоци представљају документе који су релевантни постављеном упиту, па је прецизност добијених резултата 1 или 100%.

Претрага преко пуног текста задатим упитом није дала ниједан погодак иако је упит постављен на оба писма.

Пример 2 – народна библиотека

У следећем основном претраживању поставили смо критеријум за претрагу *народна библиотека* и предвидели претрагу у свим пољима. Добили смо 11 погодака, међу којима је само 2 релевантна документа, тако да прецизност проналажења информација износи $P = \frac{2}{11} = 0,18$ или 18,2%. Приметили смо да се међу нерелевантним погоцима, осим једног оријенталног рукописа, налазе дигитални објекти о народној књижевности, а да се у метаподацима о тим објектима нигде не спомиње народна библиотека.

Напредно претраживање није дало боље резултате. Покушали смо да поставимо различите критеријуме за претрагу – кључне речи, опис, претрагу комплетног текста и нисмо добили ниједан погодак. Изабрали смо поново претрагу преко свих поља, документ у PDF формату и ограничили се на објекте на српском језику (видети слику 46). Добили смо 10 погодака. Међу погоцима није било објекта са оријенталним рукописом, а поново се само 2 објекта односи на народне библиотеке. Прецизност је у овом случају $P = \frac{2}{10} = 0,20$ или 20%.

The screenshot shows the PHAIDRA search interface. At the top, there are fields for 'Korisnik-ID:' and 'Lozinka:', along with 'Prijava', 'Pomoć', and 'Srpski' buttons. The main content area is titled 'Napredno pretraživanje'. On the left, there is a sidebar with navigation options: 'Prelistavanje', 'Osnovno pretraživanje', 'Napredno pretraživanje', 'Rezultati pretraživanja', and 'E-teze Univerziteta u Beogradu'. The main search area contains several filter rows:

- 'vrsta objekta' dropdown set to 'PDF DOCUMENT' with a red 'X' icon.
- 'pretraga svih raspoloživih polja' dropdown set to 'sadrži' with a blue dropdown arrow.
- 'srpski jezik' text input with a red 'X' icon.
- 'Jezik' dropdown set to 'srpski' with a blue dropdown arrow and a red 'X' icon.

 At the bottom of the filter area, there is a '+ Naredna "i"- relacija +' button.

Слика 46 – Резултати претраге записа у напредном претраживању у репозиторијуму PHAIDRA на упит *српски језик*

Закључили смо да претрага преко свих поља укључује и метаподатке из сегмента кључних речи означених етикетом <subject>. Такође, закључак је да ни напредно претраживање не омогућава претрагу само преко метаподатака у оквиру класификације (та опција се не може одабрати), а да приликом претраживања упита који садржи вишечлану реч, у нашем случају *народна библиотека*, претрага се врши

преко било које речи из вишечлане речи, односно фразе, у овом случају *народна*, па смо тако добили погодке који се односе на *народну књижевност*.

Претрага преко пуног текста задатим упитом ни овог пута није дала ниједан погодак иако је упит поново постављен на оба писма.

Пример 3 – Корпус савременог српског језика

У сегменту за основно претраживање одабрали смо опцију *сва поља* и поставили упит *Корпус савременог српског језика*. Добили смо одзив од 9 погодака од којих се само 1 односи на постављени упит. Прецизност овог проналажења информација је $P = \frac{1}{9} = 0,11$ или 11,1%. Анализирали смо осталих 8 погодака и закључили да се део нашег упита „*српског језика*“ јавља у кључним речима или апстрактима у овим погоцима. Дакле, индексирање се врши према било којој речи из упита што доприноси изузетно ниској прецизности. Покушали смо да претрагу извршимо тако што смо упит ставили под наводнике у циљу фразног претраживања. Добијени резултати су били идентични као и претраживањем без наводника.

Исти упит поставили смо у домену напредног претраживања, као критеријуме поставили документ у PDF формату, српски језик и у више итерација одабрали претрагу по кључној речи, опису и пуном тексту документа, али нисмо добили ниједан погодак. На задати критеријум претраге по свим пољима добили смо 9 погодака од којих само један одговара постављеном упиту. У метаподацима у осталим погоцима јављају се делови постављеног упита – *српског језика* или само *језика*. Прецизност проналажења релевантних информација је и у овом случају била 0,11 или 11,1%.

Поновили смо претраживање и у напредном претраживању одабрали да тражимо документ у PDF формату на српском језику који у наслову садржи задати упит. Добили смо 1 погодак и тај документ је релевантан, тако да је прецизност 1 или 100% (слика 47). Наш упит је био у номинативу, а у наслову публикације се задати упит јавља у генитиву – *корпуса савременог српског језика*, тако да бисмо у првом тренутку могли да констатујемо да претрага преко наслова укључује употребу морфолошких речника српског језика.

Слика 47 – Резултати претраге записа у напредном претраживању у репозиторијуму PHAIDRA на упит *корпус савременог српског језика*

Како је наш упит састављен од именице у номинативу иза које долазе три речи (придева и именице) у генитиву, поновили смо претрагу да бисмо утврдили ту претпоставку. Поред већ задатих критеријума PDF формата и српског језика, у подручје наслова уписали смо само *корпус* и добили исти релевантан документ. Још једном смо променили упит преко наслова и овог пута на упит *корпусни*, добили 2 релевантна поготка. Испитали смо још једну ситуацију – променили смо флективни облик, односно падеж свих речи из упита и употребили граматички неисправну фразу, али је она била значајна за потврђивање претпоставке – *корпусом савремени српски језик*. Као одзив смо добили 1 релевантан документ. Тиме смо потврдили претпоставку да се у овом конкретном случају упит за претрагу преко наслова може поставити преко било ког облика речи.

Променили смо упит за претрагу користећи реч која је морфолошки захтевнија јер у флективним облицима има фонолошких алтернација. У напредном претраживању изабрали смо као врсту објекта документ у PDF формату, поставили смо критеријум да наслов садржи упит *поготка* и ограничили се на документе на српском језику. Добили смо 25 погодака. Међутим, увидом у наслове пронађених докумената закључили смо да је препознат само део постављеног упита, најчешће његов почетак – *политика, подрица, полутки, података, погоном, политички, Porta, Polonia* и др. Ниједан од понуђених погодака се не односи на постављени упит, те ниједан од пронађених докумената није релевантан. Могуће је да у бази нема докумената који одговарају упиту, али се нуди листа потпуно нерелевантних одговора.

Овако конципирана претрага заправо не даје никакве резултате јер је потрага за релевантним документом практично немогућа. Неопходно је прегледати све погодке који се јављају као одзив да би се добио одговарајући документ. На претходном примеру видимо да је претрага резултирала са 0% релевантних докумената и малом прецизношћу.

Осим поменутих примера претраге, покушали смо да претраживање извршимо преко метаподатака о наставној грађи и информацијама везаним за учила и образовни процес, односно образовних метаподатака јер су они специфичност овог репозиторијума. Сви ови метаподаци су кодирани и њихов унос се врши избором одговарајућег података из понуђеног менија. На упите о врсти наставног и образовног ресурса, врсти инетрактивности, нивоу интерактивности, корисницима наставног и образовног процеса, дидактичком контексту коришћења, времену учења нисмо добили ниједан податак. Разлог за то је што, иако се пројекат којим је репозиторијум уведен на Универзитет у Београду завршио крајем 2013. године, још увек није постављена јасна стратегија, односно план за унос објеката који су везани за наставни процес и који захтевају унос ове врсте метаподатака.

Можемо закључити да проналажење информација у репозиторијуму РНАИДРА није нимало лак задатак из више разлога – иако упит поставимо на српском језику и одаберемо српски језик као језик документа, одзив се јавља у виду дигиталних објеката за документе на другим језицима, чак и на арапском језику иако поуздано знамо да не постоји могућност проширења упита на други језик. Други проблем је што се код упита састављених из више речи, претрага врши према било којој речи из упита, па је одзив у неким случајевима изузетно велики, а прецизност изузетно мала. Додатно, претрага преко наслова се врши према деловима речи из постављеног упита што даје потпуно неодговарајуће поготке. Такође, напредна претрага нуди дугу листу критеријума за избор и ограничавање претраге. Међутим, многи критеријуми се не могу одабрати (класификација, напомене, технички подаци, ентитет/лични подаци, установа, факултет, институт, детаљи о објекту, примена у настави итд.). Нарочито би била значајна претрага преко метаподатака о класификацији, посебно за кориснике каталога СОВИБ.СР који су већ обучени за ту врсту претраге. С друге стране, критеријуми се могу одабрати, али метаподаци према тим критеријумима нису унети ни о једном објекту, па претрага даје одзив у виду нула докумената.

5.2.2 Претрага записа у Дигиталној библиотеци Филолошког факултета

Дигитална библиотека Филолошког факултета омогућава претрагу преко дескриптора метаподатака и дескриптора садржаја на два нивоа – на нивоу публикације (монографске, периодичне, зборника радова, каталога или библиографије) и на нивоу чланка, односно поглавља у публикацији. У корисничком упутству³⁴⁶ за ову дигиталну библиотеку претрага према дескрипторима метаподатака назива се претрагом према атрибутима, а претрага према дескрипторима садржаја општом претрагом.

Пример 1 – биобиблиографски речник

У оквиру претраге публикација, у атрибуту који означава наслов, поставили смо упит *биобиблиографски речник* (слика 48). Као одзив на постављени упит добили смо два поготка и утврдили да су оба поготка релевантна, те да је прецизност 100%. Изабрали смо један од погодака „Биобиблиографски речник Међународног славистичког центра, 1971-2000” и утврдили да је могућа даља претрага преко комплетног текста уносом појма за претраживање.

Наслов биобиблиографски речник	Име аутора	Издавач
Наслов на страном језику	Име аутора на страном језику	Година издавања
Име приређивача / уредника	Име преводиоца	ISBN/ISSN

Општа претрага

Тражи

Слика 48 – Маска за претрагу публикација у Дигиталној библиотеци Филолошког факултета

Поновили смо речи из упита да бисмо нашли где се све тражени термин *биобиблиографски речник* појављује у тексту. Добили смо одзив од 54 погодка. Међутим, увидом у погодке, установили смо да се као резултат јављају и обе речи из упита у виду непрекидне ниске, али и засебно (слика 49).

Да бисмо утврдили да ли се претрага само према појединачним речима из упита које се могу наћи заједно или не поставили смо упит библиографски приручник. Установили смо да се као погоци јављају засебне речи из упита, те смо констатовали

³⁴⁶ Филолошки факултет Универзитета у Београду. ”Дигитална библиотека Филолошког факултета – корисничко упутство.”, <http://digifil.fil.bg.ac.rs/hlp/cvrl/Uputstvo.pdf> (приступљено 1.3.2016)

смо да овако постављен упит као резултате претраге даје један или други појам из упита или оба, те да функционише по принципу буловског оператора OR.

три стотине слависта, као и изван број научника из других области, па је то била прилика да се **биобиблиографским речником** учесника, чији су реферати били посвећени теми *Вук Караџић и његово дело у своје време и данас*, представи пресек оног дела славистичке филологије која се првенствено или у значајној мери бави српским језиком и књижевношћу.

Биобиблиографски речник који је састављен поводом тридесетогодишњице рада Међународног славистичког центра, нема јубиларни карактер. У његовој основи је замисао да се биобиблиографијама учесника међународних научних састанака покаже, колико ширина интересовања за српски језик, књижевност и културу у свету током тридесет година, толико развој наше филологије. Намера нам је да се, истовремено, потпуном библиографијом објављених радова у тридесет годишта Зборника, представе суштина и богатство научног размишљања о најважнијим темама српског језика и књижевности.

Речник је замисљен тако да на једноставан и прегледан начин библиографијама обухвати основне и важне податке о учесницима научних скупова, а да у потпуности донесе попис свих реферата објављених у зборницима Међународног славистичког центра. Стога су у **Речник** унети сви аутори и коаутори објављених радова, без обзира на то да ли постоје остали подаци о њима.

Сваку биобиблиографску јединицу чине четири дела: биографија; попис књига и важних приређених издања; списак изабраних студија; потпуни попис радова објављених у зборницима Међународног славистичког центра од првог до тридесетог годишта (бројеви непосредно иза наслова рада означавају годиште и стране). У неким биобиблиографијама сажети су други и трећи део. При обради **биобиблиографских** података у највећој мери су поштовани

Слика 49 – Избор из конкорданци на упит биобиблиографски речник у Дигиталној библиотеци Филолошког факултета

Иако су се као одзив на упите јавили различити флективни облици, није реч о употреби морфолошких речника у проналажењу информација већ о суфиксима који се надовезују на реч из упита – нпр. *библиографским речником*. Овај облик ће бити дат као резултат претраживања, али не и нпр. облици *биобиблиографског речника*, *биобиблиографском речнику*, *биобиблиографски речници* итд.

У поменутом упутству је наведено да се за фразно претраживање упит формулише тако што се испред сваке речи из упита стави знак +. Поновили смо претраживање користећи ову опцију. Овог пута смо на упит *+биобиблиографски+речник* добили 8 погодака. Сви понуђени погоци јављају се у номинативу тј. у падежу у ком је упит постављен. Да коначно утврдимо да ли постоји нека функција коју врше морфолошки речници поставили смо нови упит – *+биобиблиографског+речника* и добили један нови погодак. Тиме је, без сумње, потврђено да се при проналажењу информација не користе морфолошки речници већ само метод којим се препознаје комплетна реч из упита чак и кад има суфиксе који не постоје у упиту.

Међутим, прецизност проналажења информација је, и поред ове технолошке несавршености, прилично висока.

Пример 2 – школска библиотека

У следећем претраживању одабрали смо претрагу чланака, односно поглавља у публикацијама и у пољу атрибута (метаподатака) за кључне речи поставили упит *+школска+библиотека* јер смо желели да пронађемо оне документе у којима се јављају оба појма заједно. Добили смо 5 погодака и увидом у сваки од њих утврдили да сви погоци одговарају постављеном упиту што значи да је прецизност пронађених докумената 100%.

На овом примеру желели смо да проверимо још једну могућност претраге. Наиме, у корисничком упутству стоји да се приликом претраживања могу комбиновати упити преко атрибута и преко опште претраге и да се том приликом посматрана релација дефинише као буловски оператор AND. У оквиру атрибута, у поље за претрагу по аутору унели смо име *+Драгана+Грујић*, а у пољу општих атрибута поновили упит *+школска+библиотека*. Овог пута смо добили два погодка (слика 50). Аутор оба документа је *Драгана Грујић*, а такође у тексту оба документа јављају се све речи из упита *школска библиотека*. Тиме смо потврдили могућност комбиновања упита према различитим критеријумима претраге.

Анализирали смо, затим, појављивања траженог упита у тексту и утврдили да се погоци увек јављају у номинативу, а да нису означени погоци који су у неком другом падежу. Установили смо и да се у резултатима претраге укупан број погодака односи на целу матичну публикацију у којој је текст који смо добили као резултат претраге, те да је могуће прегледати и све остале поготке у оквиру матичне публикације добијене као одзив на тражени упит. У неким ситуацијама то је врло повољна околност јер се процес претраживања прилично скраћује и не мора се изнова вршити за сваки од докумената.

The screenshot shows a search interface with two main search panels: 'Претрага текста' (Text Search) and 'Претрага по матичној публикацији' (Search by Original Publication). Below these is an 'Општа претрага' (General Search) section with a search bar containing '+школска +библиотека' and a 'Тражи' (Search) button. The search results are displayed in a list format, showing two entries:

ДРАГАНА ГРУДИЋ
ПРИРУЧНА И СТРУЧНА ЛИТЕРАТУРА У ШКОЛСКОЈ БИБЛИОТЕЦИ у ШКОЛСКА БИБЛИОТЕКА И НАСТАВА ЈЕЗИКА И КЊИЖЕВНОСТИ, СВЕСКА 1
ДРАГАНА ГРУДИЋ
ЉУБОПИТЉИВОСТ И БАЈКОВИТОСТ у КРЕАТИВНЕ РАДИОНИЦЕ У ШКОЛСКОЈ БИБЛИОТЕЦИ, КЊИГА 1

Слика 50 – Комбиновано претраживање преко дескриптора метаподатака и дескриптора садржаја у Дигиталној библиотеци Филолошког факултета

Пример 3 – текућа библиографија

Следећу претрагу такође смо засновали на претрази чланака, односно поглавља у публикацијама и поставили упит *+текућа+библиографија* у оквиру метаподатака (атрибута) који се односе на кључне речи. Добили смо један погодак који се, несумњиво, односи на постављени упит. Прецизност је била 100%, али одзив није био висок.

Да бисмо повећали одзив и пронашли и друге документе који се тичу текуће библиографије, али можда не садрже овај појам у оквиру кључних речи или уопште немају унете кључне речи, променили смо поље за упит. Овог пута смо тражени појам унели у поље опште претраге, односно претраживали смо према дескрипторима садржаја. Добили смо 8 погодака и увидом у резултате претраживања утврдили да сви погоци садрже тражени појам. Претраживање је резултирало већим одзивом, али подједнако високом прецизношћу. Анализирали смо један од погодака у чијем се наслову не јавља тражени појам, не садржи ни кључне речи, али се у тексту тражени појам појављује– „Филолошки и етнички идентитет националне библиографије“ аутора проф. др Александре Вранеш. Информације које смо добили о појму *текућа библиографија*, а које су садржане у овом раду несумњиво говоре да је тај рад релевантан документ за наше претраживање (слика 51).

У првој половини 19. века Георгије Магарашевић и Димитрије Тирол су српску текућу библиографију обликовали по територијалном принципу, наглашавајући у њој критеријум завичајности, усмеравајући тако делатност већине библиографа, укључујући и Стојана Новаковића. Док је ретроспективна **библиографија** ненаметљиво и потпуно самостално примењивала и територијални, и национални и лингвистички критеријум, према којима се и данас обликује, **текућа библиографија** се безразложно ограничавала на територијални. Промену је унео Ђура Даничић, чија **текућа библиографија** 1858. године садржи српске књиге и нове књиге на другим језицима које се тичу Срба и то „само оне које су у Београд добављене“.

Слика 51 – Део погодака на упит *текућа библиографија* у Дигиталној библиотеци Филолошког факултета

Иако су публикације и чланци у оквиру Дигиталне библиотеке Филолошког факултета опремљени поузданим и тачним метаподацима и иако претрага преко дескриптора метаподатака даје високу прецизност пронађених информација, претрага преко дескриптора садржаја је много ефикаснија јер уз много виши одзив прецизност пронађених релевантних докумената не опада.

Једини недостатак који смо уочили је тај што процес проналажења информација није уврстио и морфолошке речнике који би омогућили претрагу преко свих флективних облика речи. Овај недостатак је делимично решен тиме што је приликом формирања упита могуће користити Клинијеву звезду ради обухватања свих потенцијалних облика речи. Међутим, тада се у конкорданцама као погоци јављају и појединачне речи из упита и у виду фразе. Додатно, постоје и случајеви када упит не можемо толико скратити Клинијевом звездом да обухватимо све потенцијалне флективне облике или је скраћивање толико велико да ће дати претрагу и оних појмова које нисмо желели. Један од примера је претрага појма *библиотека*. Да бисмо уврстили све флективне облике, Клинијеву звезду смо поставили на следећи начин – *библиот**. Овако формулисан облик упита, обухватио је и речи *библиотекар*, *библиотекарство* и *библиотечки* са њиховим различитим флективним облицима. Пошто смо желели да пронађемо информације о библиотекама, документи који садрже друга три наведена појма нису релевантни за наше претраживање, а тиме је и прецизност проналажења информација опала.

5.3 Примери претраге записа у дигиталним библиотекама применом дескриптора метаподатака и дескриптора садржаја

Претрага записа у овом одељку и анализа резултата посвећена је дигиталним библиотекама насталим у оквиру пројеката под окриљем Европеане с једне стране, а са друге стране у домаћој дигиталној библиотеци *Библиша* која садржи корпус

упарених текстова на српском и енглеском језику. Сматрали смо да су записи за дигиталне објекте у оба ресурса опремљени како метаподацима, тако и дескрипторима садржаја, па да је претрага могућа и преко метаподатака и по комплетном тексту докумената.

5.3.1 Претрага записа на порталу Еуропеана – колекције домаћих библиотека

У оквиру проналажења информација на порталу Еуропеана ограничили смо се на три колекције домаћих библиотека јер смо желели да утврдимо колико је успешно проналажење релевантних докумената који имају метаподатке на српском језику без обзира на језик публикације. Анализираћемо понаособ ситуацију у три дигиталне библиотеке– „Библиотеке Еуропеане”, „Европске новине” и „Европске колекције 1914-1918 – Портал Велики рат.” Прве две дигиталне библиотеке настале су учешћем Универзитетске библиотеке „Светозар Марковић”, а трећа учешћем Народне библиотеке Србије у пројектима Еуропеане, а на основу фондова наведених библиотека.

а. „Библиотеке Еуропеане”

Анализирали смо претрагу записа постављајући упите за које смо били сигурни да морају дати одзив будући да је Универзитетска библиотека „Светозар Марковић” приложила две колекције чија су тема кључне речи из упита.

Пример 1 – Александар Велики

У основном претраживању поставили смо упит Александар Велики и добили 213 погодака. У сегменту сужавања претраге, погледали смо постоје ли записи за документе на српском језику и утврдили да је таквих записа 24. Прегледали смо метаподатке о тим дигиталним документима. Установили смо да се међу погоцима налазе роман „Велики рат” Александра Гаталице, „Велики школски час” Александра Б. Лаковића, као и „Наш велики краљ” Александра Р. Димитријевића. Закључили смо да овако постављен упит не даје добре резултате јер проналази делове метаподатака у индексираним атрибутима за претрагу, односно део метаподатака из атрибута <title>, а део из атрибута <author>.

Променили смо захтев за претрагу. Када смо упит ставили у наводнике ”Александар Велики” добили смо одзив у виду 137 докумената. Ниједан од ових докумената није на српском језику. Желели смо да ову претрагу ограничимо на записе које је порталу Еуропеана доставила Универзитетска библиотека „Светозар Марковић” и из менија за сужавање претраге утврдили да је 123 таква документа.

Исти број докумената добили смо када смо претрагу сузили према називу колекције – колекција „Александар Велики”, односно „Alexander the Great.” Претрагу смо сузили према критеријуму језика одабравши 11 докумената на енглеском језику. Утврдили смо да је тема и ових 11 докумената на енглеском језику, као и свих 123 документа из означене колекције Александар Велики. Прецизност у пронађеним релевантним документима је 1.

Упит постављен на српском језику дао је поготке на енглеском језику због тога што се у метаподацима у атрибутима <subject> и <description> јавља текст из упита на српском језику.

Још једна позитивна функционалност коју смо уочили током ове претраге је могућност повезивања имена *Александар Велики* са другим облицима његовог имена *Alexander Magno* и *Alexander*. Ово је изузетно корисно уколико желимо да претражујемо преко ових облика имена или желимо да на једном месту објединимо записе за дигиталне објекте у којима се појављују различити облици имена једне особе, у овом случају Александра Великог. Такође, омогућена је веза ка записима за понуђене облике имена у нормативној датотеци имена VIAF, а из ове базе поново је дата веза ка запису у бази ISNI (слика 52).

ISNI:	0000 0001 2283 5816
Name:	356 Alejandro (Magno) Aleksander Makedonac'i Aleksander (Makedoneli) Aleksander Aleksander (King of Macedonia) Aleksander (ha-Gadol) Aleksander (III Wielki; (król Macedonii ;) Aleksander Macedoński Aleksander (Wielki) Aleksander Aleksander (Makedonskii) Aleksandros bar Filipos (Makdonya) Aleksanders (Lielais) Aleksandür, Makedonski Alessandro (Magno) Alessandro (re di Macedonia ; 3)

Слика 52 – Делимична листа различитих облика имена Александар Велики у бази ISNI

Занимљиво је да се у оквиру записа у бази ISNI у оквиру напомена налазе везе ка чланцима о Александру Великом на различитим језицима који су објављени на Википедији, а међу њима је и чланак на српском језику. То је још једна потврда да записи на порталу Европске библиотеке дају могућност повезивања података са другим изворима на мрежи.

Претрагу смо затим извршили у сегменту напредног претраживања. Задали смо критеријум „предмет”, односно „subject.” На наше изненађење, као одзив нисмо добили ниједан документ из поменуте колекције. Одзив је био у виду записа за 3 документа из хрватских библиотека.

Вратили смо се погоцима добијеним у оквиру основног претраживања да бисмо прегледањем метаподатака покушали да утврдимо зашто претрага преко критеријума „subject” не даје као одзив документе из посматране колекције. Прегледањем метаподатака, установили смо да је у оквиру атрибута <subject> унет податак *Александар Велики* на српском и *Alexander the Great* на енглеском језику. Како су ти подаци дати у виду веза (линкова), изабрали смо појам на српском језику *Александар Велики* и нисмо добили ниједан резултат. Избор појма на енглеском језику такође је остао без одзива. Како је атрибут <subject> у метаподацима за поједине документе једино поље где се јавља текст из постављеног упита, јасно је да је у основном претраживању тај атрибут индексан за претрагу. Ипак, нисмо у могућности да објаснимо зашто основно претраживање даје међу погоцима записе за дигиталне објекте из колекције о *Александру Великом*, али не и напредно претраживање.

Како смо навели у одељку 4.2.4, за потребе пројекта „Европске библиотеке” израђен је нови формат за унос метаподатака – ESE. Међутим, иако формат садржи бројне атрибуте и етикете за разне врсте метаподатака и иако су сви записи о претраживаној колекцији управо у овом формату, из примера претраге се види да он не даје најбоље резултате у проналажењу информација.

Одабрали смо запис за један документ из листе погодака у основном претраживању – *La vérité sur la mort d'Alexandre le Grand. La mort de Jules César* да бисмо утврдили функционалност и успешност претраге преко комплетног текста. Одабиром дигиталног објекта, добили смо везу ка репозиторијуму PHAIDRA јер је комплетан текст документа у формату PDF похрањен у овом репозиторијуму. Поставили смо различите упите за претрагу комплетног текста – „*Alexandre le Grand*”, „*Alexandre*”, „*mort*”, „*vie*”, „*Jules César*” и нисмо добили ниједан погодак. Закључак је да, иако је пројекат којим је изграђена ова дигитална библиотека (Europeana Libraries), нудио оптичко препознавање карактера (OCR), ова функционалност није искоришћена за записе за дигиталну колекцију коју је доставила Универзитетска библиотека „Светозар Марковић.”

Пример 2 – османско царство

Поставили смо следећи упит– „*османско царство*.” Овако дефинисаним упитом желели смо да као одзив добијемо и записе из колекције дигиталних оријенталних рукописа из фонда Универзитетска библиотека „Светозар Марковић.”

Као одзив на постављени упит у основном претраживању добили смо 164 погодка за документе из различитих европских библиотека. Текст тих докумената је на различитим језицима – хрватском (96 докумената), недефинисаном (24), немачком (9), енглеском (7), латинском (5), српском (5), италијанском (4), турском отоманском (4), словеначком (3) и чешком (2) језику. Погледали смо најпре поготке за документе на српском језику и установили да се код два погодка тражени појам јавља у наслову публикације – атрибут <title>, а да се у остала три случаја овај појам наведен у оквиру атрибута <subject> јер су метаподатке доставили како Народна библиотека Србије тако и хрватска Национална и свеучилишна књижница.

У другом кораку, пуну листу од 164 погодка сузили смо на записе из колекције Оријентални рукописи (Oriental manuscripts). Као одзив смо добили 4 документа. Сви они се односе на постављени упит. Језик ових докумената је турски отомански, а записи за њих су пронађени претрагом преко метаподатака на српском језику наведеним у атрибуту <description>. Међу метаподацима о овим документима нема података у атрибуту <subject> јер предметне одреднице нису ни достављене порталу Еуропеана. Како се достављање метаподатака о документима вршило извозом из електронског каталога COBIB.SR, провером у овом каталогу уочили смо да записи за наведене документе не садрже предметне одреднице.

Сваки од записа за ове дигиталне документе такође нуди везу ка дигиталном репозиторијуму PHAIDRA где се документи налазе у комплетном тексту. Будући да алати за оптичко препознавање карактера на руком писаном турском отоманском језику, у тренутку успостављања дигиталне колекције, још нису били развијени, јасно је да је немогуће претраживати преко пуног текста дигитализованих рукописа.

б. „Европске новине”

У оквиру пројекта „Европске новине” Универзитетска библиотека „Светозар Марковић” изградила је дигиталну библиотеку старих српских новина и листова са преко 400.000 дигиталних страница. Претрагу ове колекције могуће је вршити и путем

каталога Европске библиотеке и преко колекције „Претраживе дигитализоване историјске новине”³⁴⁷ на сајту Универзитетске библиотеке „Светозар Марковић.”

Како је реч о садржају новина и листова, претрага преко метаподатака даје резултате који се односе на опште метаподатке о појединим насловима новина и листова, информације о континуацији у њиховом излажењу, променама у уредништву, кратком опису садржаја сваког од наслова на српском и енглеском језику. Јасно је да тако велика колекција новинских чланака није могла бити обрађена тако да сваки чланак буде опремљен метаподацима осим коришћењем оптичке сегментације чланака што у оквиру пројекта није било предвиђено за колекцију на српском језику.³⁴⁸ Како би претрага преко метаподатака резултирала изузетно ниским одзивом у односу на постављени упит, много је ефикасније, а и занимљивије, претрагу вршити преко комплетног текста дигиталне колекције. Тврдњу ћемо поткрепити примерима које ћемо детаљно анализирати.

Пример 1 – Бранко Радичевић

Поставили смо упит у основном претраживању на порталу Европске библиотеке, добили 279 погодака и потом претраживање сузили избором картице за колекцију новина „Newspapers.” Нисмо добили ниједан документ као одзив. Потом смо у колекцији претраживих новина поставили исти упит и добили 429 погодака. У првом случају претрага је вршена преко метаподатака о новинској колекцији, а у другом преко комплетног текста новинских чланака. Додатни разлог да претрагу вршимо преко комплетног текста на порталу српских историјских новина је тај што портал Европске библиотеке нуди претрагу преко комплетног текста само за латиничне наслове из колекције новина и листова (у посматраном случају само за наслов *Srđ*).

У случају комплетног текста укључили смо целокупну колекцију новина, али је могуће извршити избор у оквиру напредне претраге и сузити критеријуме за претрагу одређеним насловима новина или листа и у одређеном периоду. Када смо у оквиру напредног претраживања поставили исти упит и претрагу ограничили на лист

³⁴⁷ Универзитетска библиотека „Светозар Марковић.” *Претраживе дигитализоване историјске новине*, <http://www.unilib.rs/istorijske-novine/pregled> (приступљено 18.2.2016)

³⁴⁸ Оптичка сегментација чланака довела би до аутоматске доделе метаподатака о сваком од чланака из новинске колекције узимајући у обзир наслов чланка, аутора, наслов часописа, годиште, годину, број и страну, односно стране на којима се чланак налази. Ова технологија је у оквиру пројекта примењена на само 2.027.967 страница новинских колекција међу којима нису новине и листови на српском језику. – извор Clemens Neudecker, Hans-Jörg Lieder and Sandra Cobel. *Europeana Newspapers – Final report*, приступљено 27.2.2016, http://www.europeana-newspapers.eu/wp-content/uploads/2015/05/Final_Report.pdf

Стражилово и период од 1885. до 1900. године, добили смо 20 погодака (слика 53). Упит смо поставили латиничним писмом иако је текст новинског корпуса на ћириличном писму јер писмо није од значаја приликом проналажења информација на овом порталу.

Претраживе дигитализоване историјске новине

Претрага Напредна претрага Преглед

branko radičević

10 резултата по страници Поређај по скору Стражилово

Од: 1885 До: 1900

Наћи

Резултати 1-10 од 20 пронађених. Претрага трајала 51 милисекунду. Следећи >>

Стражилово, страна 14 31. 5. 1892.

-кз 350 Стјић, рођсној **Радичевић** "ј јер кућа **Радичевића** у ВојБевцима ојеше тада већ ева обамрла у ... Радичевићима одакле еу и шта еу, отуда у њега иријенис и тестапелта и посмртног инвентара Тодора **Радичевића** ... је у прози, 13 а Р>уко10 И даис стоји у Бол.снцмгаа л кућа и иде **Радичевића**, али то иијевисекрв онијех **Радичевића**. 11 Исиор. нисжа Вука Отеф. Карацића Јустипу Мпхајлрвићу. „Јавор" 1887. У инсму од 13 ... , коју му је испјевао год. 1844. под патписом „Моје сунцс." 11 екада. је —• вели **Бранко** — имао још једпо

Стражилово, страна 14 15. 1. 1887.

начисто бити. *ИВШВВ Х Г Л А О Н И К. („Где станак мој.. и **Бранко Радичевић**.) Позната ће бити нашим" Али како је преведена на српски, и да ју је превео баш сам **Бранко Радичевић** пре 35 година, то ... , зашто да баш није слободно српски певати, он се лукаво насмеши, па их упуту Бранку **Радичевићу**, да им ову ... вечери нађоше Бранка код куће, где кува чај и замоле га, да преведе чешку химну. **Бранко** их замоли, да

Стражилово, страна 4 11. 2. 1888.

. Него **Бранко Радичевић** дубоко је заталасео најврелије осећаје срца и душе српског народа својим лаким ... Бранка: — Хоће ли ко говорити штогод том приликом? — То је бреме нало на мене, — одговори **Бранко**. — Аха ... подухватио у иајбољој вољи. — Немојте се плашити, господине **Бранко**, — рече Даринка, као да јој жао, ако је ... и Јован, а и **Бранко** пристао уз њих, ... упели се, па доказују Јелени, како митроносни песник омладине ... **Радичевића**. Необични дар, неуморни труд и сувремена ученост и Рајића и Мушичког дуго ће остати узором

Стражилово, страна 2 21. 7. 1888.

Бранко Радичевић, онет Карловачки ћак, у својим песмама као нико ни пре ни носле њега. Е, па спомену тога ... у сребрним и златним кивотима у велелепне храмове. **Бранко Радичевић** беше, наравно, само сиромашан

Стражилово, страна 3 25. 2. 1888.

Даринка, Јелка, **Бранко**, Јован, а уз њих трчка и мала Зорка. И њено срдашце куца необично живо, јер види ... само толико радује том Јелена и Р П С К О Г А Ж И В О Т А. ДЕО. Даринка, **Бранко** и Јован, него им ... речи већ о прослави Бранка **Радичевића** у Брајковцу. Зорка неје дала нн о чему другом да се говори, него

Слика 53 – Пример претраге новинске колекције упитом *Бранко Радичевић* са делом конкорданци

У резултатима је наведен наслов листа, односно новина, датум када су издате, исписан је део пуног текста, а означен је тражени појам. Изабрали смо један од погодака (слика 54).

. године Бранку **Радичевићу**", њему с десне „Клек“, с леве „Пљешевица“; а са северне у средини „Велебит ... и **Бранко** није питао, које и колико граница деле Сриство тамо и овамо. Он је певао јединству свог народа, те

Слика 54 – Пример једног резултата претраге новинске колекције на упит *Бранко Радичевић*

Ако погледамо резултате одзива на постављени упит, запазићемо да су као погоци означени различити морфолошки облици (*Бранко Радичевић*, *Радичевића*, *Радичевићу*). Да бисмо проналажењем информација добили овакве погодке, неопходно је да на текст корпуса буду примењени морфолошки речници. Међутим, приметимо да је целокупан упит означен само у оним погоцима у којима се тражени појам јавља у номинативу, а да су елементи упита означени и у случајевима када не стоје непосредно један уз други.

У следећем кораку повезали смо се са пуним текстом конкретног чланка у коме се јавља тражени појам (слика 55). Чланак који смо добили претрагом објављен је у рубрици *Смесице*, а у њему се говори о споменику Бранку Радичевићу и месту где је сахрањен, као и о неким елементима из биографије овог славног писца. На слици текста чланка видимо да су појмови које смо тражили означени црвеном бојом, те да су засебно пронађени елементи *Радичевићу* и *Бранко*, али не и одговор на цео упит *Бранку Радичевићу* што би уз квалитетну примену морфолошких речника морало да се догоди.

Закључили смо да на корпус текстова нису примењени морфолошки речници већ се у проналажењу информација јавља проналажење префикса – *Радичевић*, што затим укључује и све суфиксе које се појављују уз тражени појам *Радичевић-у*, *Радичевић-а*. Потврду за ову тврдњу налазимо и у примеру када различити флективни облици имена *Бранку*, *Бранка* нису препознати јер не садрже цео префикс – *Бранко*.

Свакако да је претрага преко комплетног текста изузетно значајна и кориснику даје много веће могућности претраге, те резултира већим одзивом и прецизношћу у односу на претрагу преко метаподатака, али је, такође, пре свега значајно да корисник пронађе све информације које одговарају постављеном упиту.

ра“, њој с десне „Ловћен“, с леве „Клек“; са западне стране је у среди камен с натписом „С Вршачке Куле 1883. године Бранку Радичевићу“, њему с десне „Клек“, с леве „Пљешевица“; а са северне у средини „Велебит“, њему с десна „Пљешевица“, с лева „Авала“. У дну последње степенице, а на челу споменику, опет с источне стране, одмах на уласку кроз гвоздену ограду, што је подигнута око споменика, узидана је обновљена плоча с песникове гроба из Бечког још гробља, с натписом, који је био и на старој плочи, што је Бранку још покојни му отац на гроб ставио, (а данас се чува као спомен у музеју народном у Београду, камо је послао бечки одбор 1883.):

Алекса Радичевић

рођен 1824. а умре 1853.

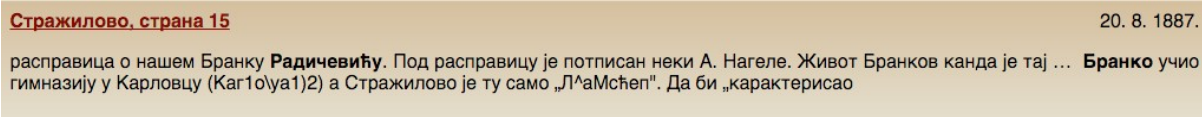
Много тео, много започео,
Час умрли њега је помео.“

Испод тога је дао урезати бечки одбор: „Пренешен (!) на Стражилово у Карловцима 10. јула 1883.“ Споменик је скроман, али према месту, на ком је, величанствен. Израђен је по замисли самога одбора, а по нацрту Владимира Шајковића и Светозара Ивачковића, архитекта у Београду, који је градио и велику папчевачку цркву, као и капелу Харишеве породице на земунском гробљу. Испред споменика је прокрчен брег, да се одонуд може догледати далеко у Бачку и Банат преко Дунава и Тисе, па и онамо преко Банстола до Авале. Одбор би добро урадио, када би на згодном месту подигао још и таблу за оријентирање на све стране, камо се даде догледати, као што то има свугде на високим местима, па би посетиоце занимало, да прате отуд погледом, где лежи какво место. Уз вис води до гроба поширока стаза у кривуд, а украј ње је на неколико места клупа за одмарање, као што их има и на самом вису. — Споменик тај је први споменик, што је подигнут прилозима српског народа из свију крајева без разлике, камо који крај спада. Али и Бранко није питао, које и колико граница деле Српство тамо и овамо. Он је певао јединству свог народа, те и народ, испунивши му жељу преносом, као видљиво знамење тог јединства диже такав спомен на гробу свог љубимца!

Слика 55 – Комплетан текст чланка са препознатим речима из упита

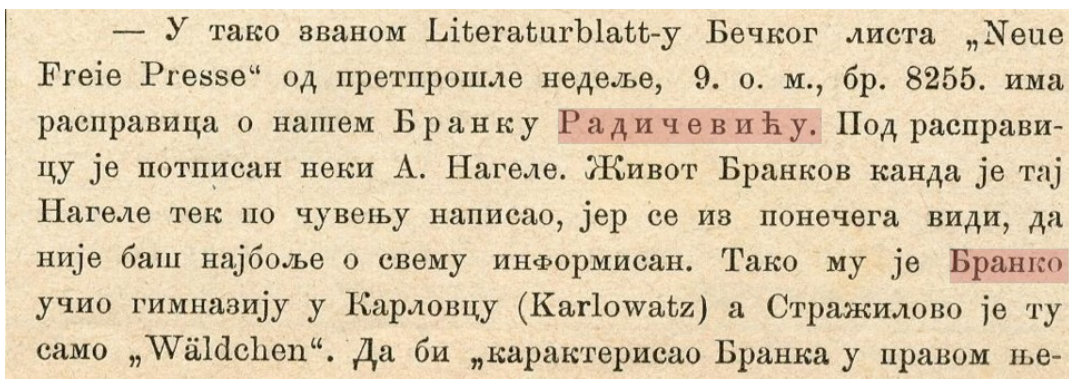
Избором једног другог погодка утврдили смо да оптичко препознавање карактера није увек довољно функционално. Карактери на ћириличном писму су исправно

препознати, али пошто није извршен OCR за латинично писмо, остале су непознате две речи у примеру на слици 56:



Слика 56 – Грешке у OCR-у – непознати латинични карактери у ћириличном тексту

Када смо погледали дигиталну копију оригиналног текста, утврдили смо да нису препознате следеће речи на немачком језику: *Karlowatz* и *Wäldchen*.



Слика 57 – Оригинални текст са латиничним писмом у ћириличном тексту

У контексту исправног оптичког препознавања карактера, потребно је навести да је у коначном извештају о резултатима пројекта Европске новине наведено да је исправност препознатих карактера за српски језик на ћирилици 76,1%³⁴⁹. Како је близу 25% неправилно препознатих или непознатих карактера, постоји могућност да нека претрага не даје резултате и из овог разлога. И у случају када карактери нису исправно препознати, примена морфолошких речника бикла би од користи за квалитетније проналажење информација.

У истом извештају утврдили смо да је као метод класификације и представљања докумената на начин разумљив рачунару коришћена техника класификације на нивоу речи тзв. „врећа речи“ (енг. “bag of words”) која се односи на то да се документ представља као вектор речи које се појављују у њему, а где сваки елемент вектора садржи информацију о фреквенцији дате речи у документу.³⁵⁰ Управо је ово разлог да проналажење информација из документа даје одзив без обзира на редослед речи у

³⁴⁹ Clemens Neudecker, Hans-Jörg Lieder and Sandra Cobel. *Europeana Newspapers – Final report*, http://www.europeana-newspapers.eu/wp-content/uploads/2015/05/Final_Report.pdf (приступљено 27.2.2016)

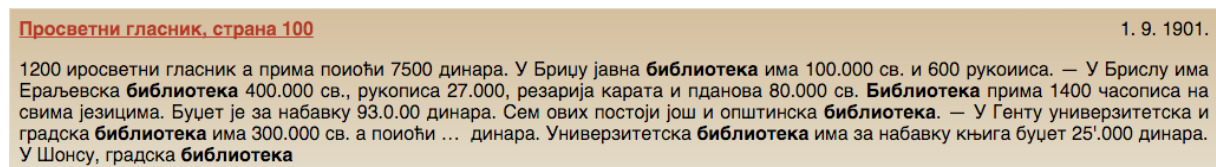
³⁵⁰ Јелена Граовац. „Класификација текста заснована на српском wordnet-у.“ *Инфотека* 14,2(2013):4, http://infoteka.bg.ac.rs/pdf/Srp/2013-2/INFOTHECA_XIV_2_2013_2-17.pdf (приступљено 3.3.2016)

постављеном упиту и њихову повезаност у тексту. У потоњим примерима претраге новинске колекције видећемо да је то и случај.

Пример 2 – народна библиотека

У претраживачу дигитализованих новина у комплетном тексту поставили смо упит *народна библиотека* и добили одзив у виду 965 погодака. Увидом у листу погодака, утврдили смо да се у исписаним деловима текста (конкорданцама) јавља исправно препознат појам *народна библиотека*.

Међутим, приметили смо један интересантан пример, приказан на слици 58:



Слика 58 – Пример претраге новинске колекције где је пронађена само једна реч из упита

Као пронађени појам означена је само реч *библиотека*, али не и *народна*. Очекивали смо да ће у свим погоцима претрага резултирати пронађеним информацијама које садрже сваку реч из упита у виду непрекидне ниске. Овај пример показује да то није случај. Чак се међу конкорданцама нигде не појављује реч *народна*.

Што је још занимљивије појављују се термини *јавна библиотека*, *Ераљевска библиотека* (неисправно препознат карактер – *Краљевска библиотека*) *општинска библиотека*, *градска библиотека*, *Универзитетска библиотека*. Иако су термини *општинска библиотека*, *јавна библиотека*, па и *градска библиотека* данас условно синоними за *народну библиотеку*, ипак овај погодак није одговорио на постављени упит јер смо очекивали да ћемо као одзив добити с једне стране документе који говоре о Народној библиотеци Србије и другим националним библиотекама јер се у то време појам *народна библиотека* односио на националну библиотеку код нас и у свету, а с друге стране и документе о општинским, градским и јавним библиотекама.

Ипак, у наведеном примеру није случај да је систем за проналажење информација препознао синониме. Систем није опремљен таквом технологијом. Реч је, заправо, о ситуацији да постављени упит функционише као Буловски оператор ИЛИ што значи да се као резултат претраге појављују речи *народна* или *библиотека*.

Да се термин *народна библиотека* односи и на националну библиотеку видимо из следећих погодака (слика 59) где се јављају *Грчка Народна библиотека* и наша *Народна библиотека*:

Општинске новине, страна 53

1. 6. 1940.

7* Солун и Атина 563 Грчка **Народна библиотека** споменули у почетку као примере неокласичне ... , младог ерудита г. Нициаса Вертиса. **Народна библиотека** је највећа **библиотека** у целој Грчкој. Укупан број књига износи пола милиона, дневно прође кроз библиотеку 100—300 читалаца. По свом саставу **Библиотека** Камбуроглу-а. Архив претставља ванредан извор за упознавање новије грчке историје. **Народна библиотека** ... скромно; тако и монаси, који су увек задубљени у диспуте. **БИБЛИОТЕКА** ГЕНАДИОН У подножју Ликабета, — с оне

Просветни гласник, страна 23

1. 2. 1901.

. Ако је више приновљених књига укoricено уједно, а **Народна Библиотека** нема ни једне од њих, онда ће све ... **Народна Библиотека** има само неке од њих, онда ће се и така укoricена свеска сматрати као приновљена, те ће ... посебним нумерама књиге, које **Народна Библиотека** већ има, међусобно, а у азбучном и стручном каталогу ... укoricено уједно, а **Народна Библиотека** има свако од њих и посебно, онда се така укoricена свеска ипак сматра

Слика 59 – Пример погодака у новинској колекцији који се односе на националне библиотеке

На основу следећег примера констатовали смо да, осим што се не проналазе све речи из постављеног упита у виду непрекидне ниске, ова претрага није дала ни резултате у препознавању различитих морфолошких облика постављеног упита (слика 60).

Просветни гласник, страна 22

1. 6. 1903.

ИЗВЕШТАЈ НАРОДНЕ БИБЛИОТЕКЕ У БЕОГРАДУ За 1902. годину Госиодгту Министру Просвете и Црквених Лослова. Пошто је **Народна Библиотека** двадесет година живела по одредбама свога старог закона, увидело се најзад да, поред свега што су за то време били као библиотекарски на челу ове установе угледни представници ... лањском је извештају казано, да је **Народна Библиотека** имала на дан 1. јануара 1902. године: 34.130 дела ... старе незгоде од којих **Библиотека** трни а које раније можда нису биле ни уочене. Господин Министар

Слика 60 – Пример непрепознавања различитих морфолошких облика у претрази новинске колекције

Препознат је облик Народна библиотека у два случаја, али не и облик Народне библиотеке са почетка текста.

Да бисмо са тачношћу утврдили да ли се као одзив на постављени упит јављају погоци са различитим морфолошким облицима, сузили смо претрагу. Тражени појам смо ограничили само на корпус текстова из „Малих новина.” На постављени упит народна библиотека добили смо 35 погодака. Сви пронађени термини били су у номинативу, односно били су идентични постављеном упиту. Променили смо упит. На упит *народне библиотеке* добили смо нових 16 погодака, на упит *народних библиотека* 3 погодка, а на упит *народним библиотекама* 1 погодак.

Покушали смо да приликом постављања упита користимо Клинијеву звезду не бисмо ли обухватили све потенцијалне морфолошке облике вишечлане речи у упиту. Нови упит је изгледао овако: народн* библиот*. Нисмо добили ниједан погодак јер скраћивање појмова Клинијевом звездом није предвиђено.

Ипак, без обзира на очигледна ограничења и непрецизност, као и немогућност проналажења искључиво оних докумената који садрже све речи из задатог упита као непрекидну ниску, ова колекција докумената нуди много квалитетније проналажење информација у односу на документе који су опремљени само метаподацима.

Пример 3 – Церска битка

За претрагу овим упитом смо се определили да бисмо упоредили проналажење информација у оквиру дигитализованих историјских новина у односу на проналажење информација на тематском порталу Велики рат.

На постављени упит добили смо 36 погодака у оквиру целокупне колекције новина и сваки од погодака се односи на задати упит. Прецизност је, дакле, изузетна.

Пошто се у литератури јавља и назив *Битка на Церу*, нови упит смо тако дефинисали и добили више пронађених докумената – 69 погодака. Увидом у листу погодака, утврдили смо да 3 документа не одговарају постављеном упиту (из ниске речи из упита у ова три случаја препознат је само предлог *на*).

в. „Европске колекције 1914-1918 – Портал Велики рат”

Иако је реч о дигиталној библиотеци која је изграђена у оквиру пројекта Еуропеане, проналажење информација могуће је вршити искључиво са портала Велики рат који је део дигиталних збирки Народне библиотеке Србије. Како је реч о тематском порталу који се односи на Први светски рат, упите смо дефинисали тако да су у складу са овом темом. За претраживање докумената понуђено је основно и детаљно претраживање. Детаљно претраживање нуди могућност сужавања резултата претраживања избором одређених критеријума – атрибутима из формата Даблинско језгро или избором колекције у којој се документ налази.

Пример 1 – Церска битка

Поставили смо упит са критеријумом „кључна реч” и добили 14 докумената као одзив. Провером метаподатака о тим документима установили смо да је само 6 релевантних докумената што чини да је прецизност пронађених информација $P = \frac{6}{14} = 0,42$ или 42%. Погледали смо детаљне метаподатке о сваком од понуђених погодака. Свих осам нерелевантних погодака садрже међу метаподацима реч битка, седам од њих се односе на друге битке из Првог светског рата, али не и на Церску битку док се један нерелевантан документ тиче прославе годишњице Косовске битке (текст је на енглеском језику).

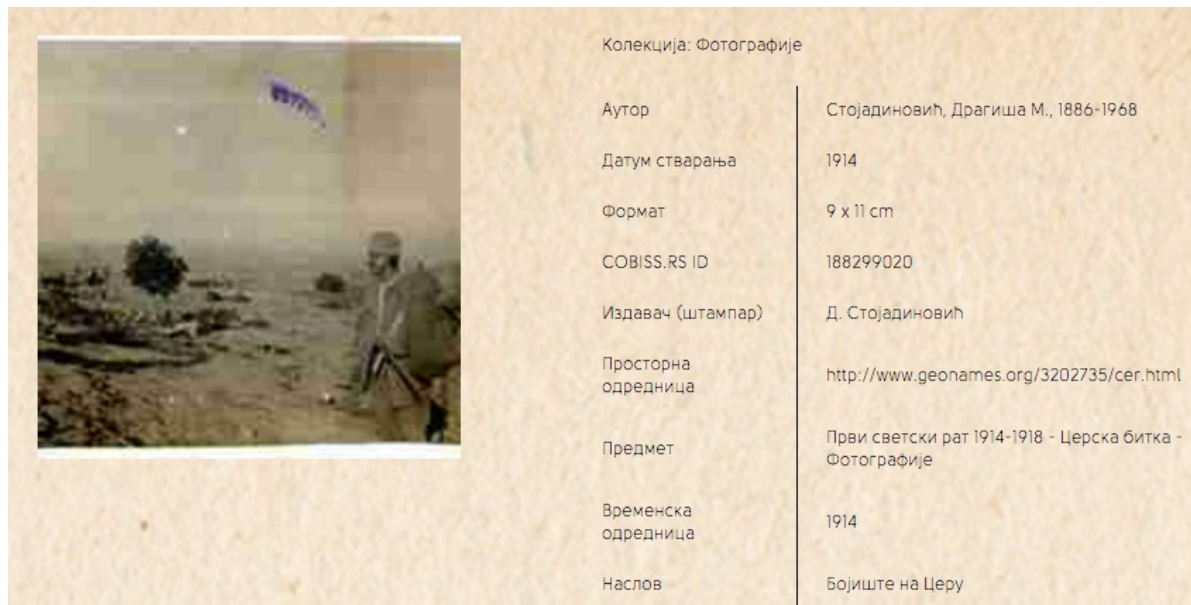
Покушали смо, да као и у претходном случају променимо упит и пронађемо документе који ће одговорити на упит *Битка на Церу*. Дobili смо идентичне погодке као и у случају упита *Церска битка*. Разлог за то је што се у оба упита садржане речи *битка* и *Цер*.

Детаљна претрага нуди могућност сужавања резултата према садржају одређених поља. Када смо претрагу сузили само на предмет документа, поново смо добили 14 погодака иако смо поуздано утврдили да се само 6 њих односи на постављени упит. Индексирање преко предмета такође је извршено тако да претражује било коју реч из упита, а упит није могуће поставити тако да се као резултат претраге добију само они документи који садрже све речи из упита.

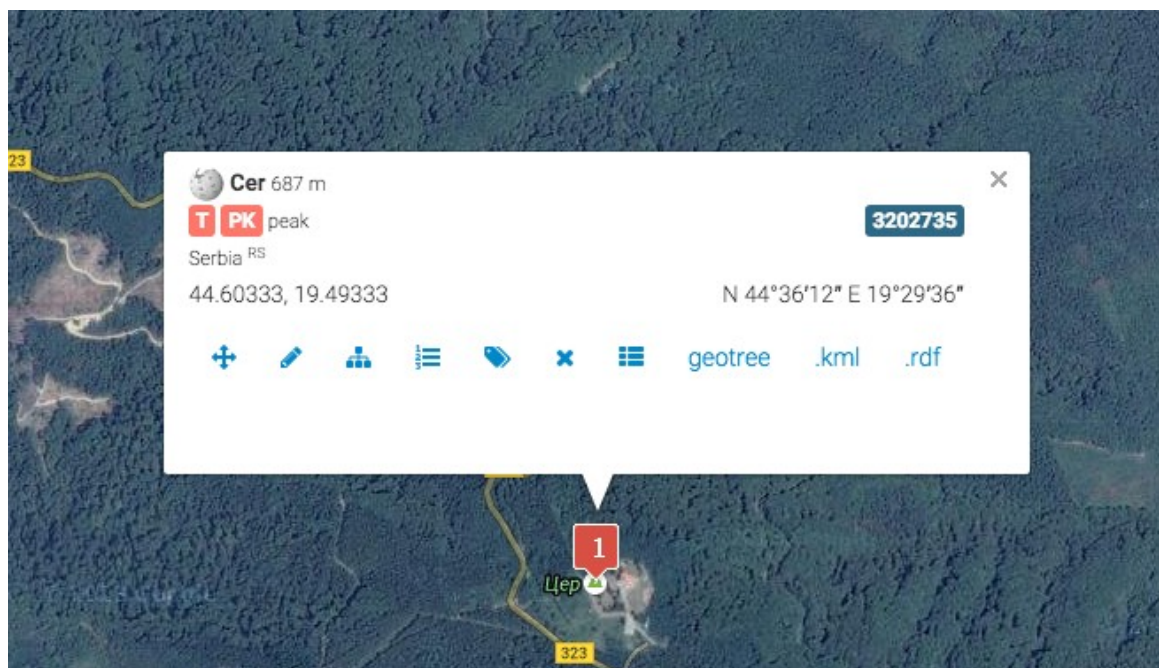
Увидом у опције за ограничавање резултата претраге увидели смо да претрага преко пуног текста није понуђена. Остали смо ускраћени за информацију да ли је то само тренутна ситуација будући да је пројекат у току или је у питању коначно решење.

Претрагу на овом порталу могуће је ограничити и према колекцијама – књиге, периодика, слике (разгледнице, фотографије, албуми), плакати, разно (споменици, заставе), рукописи, картографска грађа и индекс аутора. Како смо у претходном одељку претраживали новинску колекцију и на порталу Велики рат смо одабрали новинску колекцију, али на упит *Церска битка* нисмо добили никакав одзив. Разлог за то је што је у новинској колекцији могуће вршити претрагу само према метаподацима о одређеном наслову новина, али не и по његовом садржају.

Међутим, како сви пројекти развијани у оквиру Еуропеане пропагирају позитивне особине семантичког веба и омогућавају везу записа за дигиталне објекте са Отвореним повезаним подацима, тако се и у записима за дигиталне објекте на порталу Велики рат могу видети везе ка бази GeoNames. Навешћемо један такав пример из погодака које смо добили на упит *Церска битка* (слика 61). У метаподацима за просторну одредницу (атрибут <dc:coverage>) видимо да је унета веза ка бази GeoNames за ороним *Цер* (слика 62).



Слика 61 – Пример записа са метаподацима на порталу Велики рат



Слика 62 – Веза на бази GeoNames из записа на порталу Велики рат

Пример 2 – Солунски фронт

Претпоставили смо да ћемо постављањем овог упита као одзив добити прилично велики број докумената, па смо сузили критеријум за претрагу на „просторну одредницу”, односно на атрибут <dc:coverage>. Овим упитом и критеријумом пронашли смо 90 докумената. Погледали смо целокупну листу погодака и припадајуће метаподатке за сваки од њих и утврдили да се 16 пронађених докумената

не односи на постављени упит. Прецизност пронађених релевантних докумената је, дакле, $P = \frac{74}{90} = 0,82$ или 82%. Утврдили смо да нерелевантни документи немају унете метаподатке у оквиру атрибута <dc:coveradge> нити се међу осталим метаподацима налази појам *Солунски фронт*. Једина веза са постављеним упитом јесте реч *фронт* која се на српском или енглеском језику јавља у метаподацима у 13 случајева. У три случаја нисмо нашли никакву везу са постављеним упитом јер метаподаци не садрже ни појам *Солун (солунски)* ни појам *фронт*.

Осим неадекватних резултата претраге које смо добили, замерка је и та што се након обављеног претраживања не појављује бројач који показује колико је докумената пронађено након постављеног упита, па се записи морају листати да би се утврдио тачан број погодака. Такође, овако велика колекција која не садржи само сликовну грађу већ и велику количину текстуалне грађе морала би да буде опремљена дескрипторима садржаја да би понудила квалитетно проналажење информација.

5.3.2 Претрага записа у дигиталној библиотеци Библиша

Дигитална библиотека *Библиша* нуди претрагу преко дескриптора метаподатака и дескриптора садржаја. У другом случају користе се лексички ресурси који омогућавају морфолошка проширења упита и проширење упита на други језик. Такође, од значаја је и то што и овај систем омогућава да се упит на српском језику може поставити на оба писма – ћириличном и латиничном. Особине дигиталне библиотеке *Библиша* детаљно су описане у одељку 4.4, а у овом одељку ћемо анализирати ефикасност проналажења информација у односу на постављени упит, најпре претрагом преко метаподатака, а потом и преко комплетног текста докумената за сваки од примера.

Пример 1 – универзитетска библиотека

У дигиталној библиотеци *Библиша*, одабрали смо претрагу преко метаподатака, поставили упит, претрагу ограничили на српски језик, на метаподатке о чланцима из часописа „Инфотека” и на то да се тражени појам из упита појављује као кључна реч (слика 63).

Слика 63 – Претрага преко метаподатака у дигиталној библиотеци Библиша

Као одзив на постављени упит добили смо један погодак и утврдили да метаподаци о пронађеном документу садрже појам *универзитетска библиотека*. Ради јаснијег приказа података, као илустрацију добијеног погодка одабрали смо почетак исписа у ТМХ (слика 64) у виду паралелног текста на енглеском и српском језику.

Paraleni prevod: Akhund, 2011, vol. XII:1, ID: 1.2011.1.1	
English- (414)	Srpski - (414)
n1 : The work of the Carnegie Endowment in the Balkans after World War One: The University Library of Belgrade, 1919-1926	n1 : Delatnost Kamegijevih zadužbina na Balkanu posle Prvog svetskog rata: Univerzitetska biblioteka u Beogradu, 1919-1926
n2 : Nadine Akhund, Columbia University, New York	n2 : Nadin Akund, Kolumbija univerzitet, njujork
n3 : Keywords: Carnegie Endowment for International Peace, University Library Belgrade, building	n3 : Ključne reči: Kamegijeva fondacija za mir, Univerzitetska biblioteka u Beogradu, izgradnja

Слика 64 – Пример претраге кључном речју *универзитетска библиотека* у дигиталној библиотеци Библиша

Исти упит смо поставили у сегменту претраге пуног текста докумената. Означили смо српски језик, корпус текстова часописа „Инфотека” и започели претраживање. Утврдили смо да се тражени појам налази у лексичким ресурсима wordnet и Библиотекарски термиолошки речник. Затим смо одабрали морфолошко проширење упита и испис конкорданци где се види да се тражени појам јавља 73 пута у одабраном корпусу текстова. Увидом у листу конкорданци приметили смо да су препознати различити морфолошки облици вишечлане речи из упита (*универзитетска библиотека, универзитетске библиотеке, универзитетску библиотеку*), као и да су, осим погодка на српском, у паралелном тексту означени и погоци на енглеском језику (*university library, university libraries, academic libraries*),

те да се све конкорданце односе на постављени упит што значи да је прецизност пронађених релевантних докумената 1 или 100% (слика 65).

n105 : <i>University Library</i> (http://www.unilib.bg.ac.yu), as one of the largest research libraries on Balkans, provides information for scientists and faculty staff, and accordingly entire academic population students, university teachers and researchers.	n105 : <i>Univerzitetaska biblioteka</i> (http://www.unilib.bg.ac.yu) kao jedna od najvećih naučnih biblioteka na Balkanu ima ulogu da pruži informacije naučnom i nastavnom osoblju, odnosno celokupno akademskoj populaciji -studentima, univerzitetaskim profesorima i istraživačima.
n107 : The <i>University Library</i> is, by virtue of its domain, master library for all <i>academic libraries</i> in Serbia and depository for master (M.A. and M.S) and doctoral (PhD) dissertations.	n107 : Po svojoj funkciji <i>Univerzitetaska biblioteka</i> je matična za sve visokoškolske biblioteke u Srbiji i deponitna za magistarske i doktorske radove.
n129 : due to the fact that automated book lending was initialized on 8th January 2003 in <i>University Library</i> , having been the first of the kind in COBISS.SR system.	n129 : Razlog za njegovo pokretanje jeste i činjenica da je u <i>Univerzitetaskoj biblioteci</i> , prvoj u sistemu COBISS. SR, pokrenuta automatizovana pozajmica (8. januara 2003.).
n155 : There is a lot of work to be done on retrospective cataloguing at the <i>University Library</i> .	n155 : <i>Univerzitetasku biblioteku</i> čeka još dosta posla na retrospektivnoj katalogizaciji.

Слика 65 – Избор из конкорданци добијених на упит *универзитетска библиотека у дигиталној библиотеци Библиша*

Пример 2 – Корпус савременог српског језика

Упит смо најпре поставили у сегменту претраге преко метаподатака. Одабрали смо српски језик, колекцију текстова часописа „Инфотека” и претрагу према речима из наслова, те смо као одзив добили један документ. Утврдили смо да је тај документ релевантан постављеном упиту и да је пун наслов чланка „Анотација Корпуса савременог српског језика.”

У следећој претрази одабрали смо претрагу преко комплетног текста. Појам из упита се најпре није јавио ни у једном од три означена речника, а променом колекције текстова где смо одабрали целокупну колеквицију, појавио се у два речника на српском језику – GeoIISSTerm и RudOnto. Ипак, и у првом случају постављени упит дао је као одзив 10 конкорданци из три различита документа. Све конкорданце су релевантне за наше претраживање, тако да је прецизност 1. Као и у претходном примеру, у конкорданцама су означени различити морфолошки облици вишечлане речи из упита (*Корпус савременог српског језика*, *Корпуса савременог српског језика*), а јављају се и погоци на енглеском језику (*Corpus of Contemporary Serbian*). Приметили смо да се поред пуног назива јавља и скраћеница *СрпКор*, те смо утврдили да тај облик треба такође унети у лексичке ресурсе и повезати га са пуним називом ради побољшања претраге.

У односу на претрагу преко метаподатака, као резултат претраживања добили смо један идентичан и два нова текста. Како су сви документи релевантни, закључићемо да је претрага преко комплетног текста дала боље резултате у смислу одзива, док је прецизност иста.

Пример 3 – мултимедијални документ

У следећем примеру претраге преко метаподатака анализирали смо проналажење информација постављањем критеријума српски језик, колекција текстова часописа „Инфотека”, речи из наслова у првом претраживању, а кључне речи у другом претраживању. У првом случају добили смо три погодка, а у другом два. Утврдили смо да трећи текст уопште не садржи кључне речи, па је то разлог што се није нашао међу погоцима. У оба претраживања сви пронађени документи су релевантни, те је прецизност 1 или 100%.

Приступили смо анализи претраге комплетног текста и поставили упит који је пронађен у лексичким ресурсима GeolISSTerm и RudOnto, а као резултат смо добили и 67 конкорданци из 4 различита документа. Три пронађена документа претходно смо пронашли и претрагом преко метаподатака, док се променом начина претраживања појавио један нови документ. У конкорданцама се јављају различити морфолошки облици вишечлане речи из упита на слици 66 (*мултимедијални документ, мултимедијалног документа, мултимедијалном документу, мултимедијалне документе, мултимедијалним документима*), као и еквивалентни погоци на енглеском језику (*multimedia document, multimedia documents*).

n183 : Садржај **мултимедијалног документа** употпуњују и критике следећих позоришних представа објављене у различитим новинама и часописима; Афера Лјилјак, Бела кафа, Чарапа од сто петљи, Чарлама збогом, Капа доле, Комунистички рај, Крмећи кас, Лјубинко и Десанка, Мрављи метеж, Пепелјуга, Развојни пут Боре Шнајдера, Сабља димискија, Смртоносна мотористика, Свети ђаво Распућин и Утва птица златокрила.

n184 : Будући да се сав прикупљени материјал није могао наћи на **мултимедијалном документу**, студенти Математичког факултета су направили и мултимедијалну базу у коју су уврштене бројне фотографије, плакати, исећи из новина, разни критички текстови, делови интервјуа, а која је доступна преко интернета (<http://multimedia.math.rs/index.html>).

n186 : **Мултимедијални документ**, као студентски предмет-пројекат, новина је у образовном процесу на Катедри за библиотекарство и информатику.

n187 : Осим што је, кроз различите евиденције, праћена активност студената у току израде **мултимедијалног документа**, током семестра обратили смо пажњу и на друге квалитативне показатеље;

Слика 66 – Избор из конкорданци добијених на упит мултимедијални документ у дигиталној библиотеци Библиша

Можемо закључити да претрага преко метаподатака у овој дигиталној библиотеци даје добре резултате и да је најефикаснија ако желимо да пронађемо радове једног аутора, док је претрага преко комплетног текста препоручљива уколико претражујемо кључном речју или тражимо информације о одређеном појму. У свим наведеним примерима, утврдили смо да су се конкорданце без изузетка односиле на постављени упит, те да је прецизност претраживања изузетна.

Након претраге записа преко дескриптора метаподатака у свим анализираним ресурсима закључујемо да особе које раде на додели метаподатка треба да буду изузето добро обучене да би могле да претпоставе на које све начине корисник може поставити упит. У складу са тим је и потреба да се у запис о дигиталном или аналогном објекту укључи што више метаподатака и тако повећа одзив и прецизност у проналажењу информација.

Анализом претраге преко дескриптора садржаја закључили смо да корпус докумената који се претражују преко комплетног текста мора бити опремљен квалитетним лексичким ресурсима и морфолошким речницима да би проналажење информација дало релевантне резултате. Постојање огромних текстуалних колекција које није могуће претражити преко комплетног текста је неисплативо и неефикасно јер претрага преко метаподатака не даје прецизне резултате. Изостанак лексичких алата чини да се упит мора поставити више пута, као и да се морају мењати критеријуми за претраживање.

На примерима претраге у анализираним ресурсима утврдили смо да дигитална библиотека која у проналажењу информација примењује морфолошке речнике нуди неупоредиво ефикасније и прецизније проналажење информација од оне у којој то није случај. Такође, треба имати у виду да дигитализација једне колекције није завршена нити је дигитална библиотека креирана све док се на дигиталном тексту не примене лексички ресурси, изврше корекције пронађених грешака, а потом употребе и алати за квалитетну претрагу.

6 Надградња и допуна лексичких ресурса српског језика

Један од циљева овог истраживања је надградња и допуна лексичких ресурса српског језика са посебним фокусом на терминологији из области библиотекарства и информатике. У поглављу 4.4 наведени су лексички ресурси који се користе у овом истраживању, а резултат истраживања биће надградња и допуна терминолошког електронског речника српског језика из библиотекарства и информатике и морфолошког електронског речника српског језика.

6.1 Терминологија

Пошто је терминологија, као наука која се бави проучавањем термина и њихове употребе, једна од веома важних области у обради природних језика, потребно је најпре дефинисати појам и дати историјски осврт развоја ове области.

6.1.1 Историјски осврт

Претеча појма терминологија је појам номенклатура (од лат. *nomenclature* са значењем *називати неким именом*; *nomen* – име, *clature* – називати) који се користио од 16. века, најпре у енглеском и француском језику³⁵¹ као назив за систем који означава посебне категорије речи, односно именује појмове,³⁵² или као „израз који тачно означава нешто из области науке или уметности, стручан израз“,³⁵³ односно „1. укупност назива потребних једном знанственом подручју, некој струци или дјелатности; збирка термина; називље; 2. знаност о тим терминима“,³⁵⁴ а различит од оних имена и појмова који се јављају у општим речницима.

У 18. и 19. веку научници из природних наука били су носиоци рада на попису појмова из својих специфичних области, али и на успостављању веза између појмова и концепата. На подручју номенклатуре, истиче се рад Карла фон Линеа (Carl von Linné 1707-1778) који је своје дело „Систем природе“ (лат. *Systema naturae*) објављено 1735. године базирао на систематици и класификацији биљака. Данас се Лине сматра оцем таксономије (класификационе науке чији је основни принцип да су објекти и феномени подељени у класе, а они у поткласе итд.), а познат је и по томе што је увео биномијалну номенклатуру (именовање врсте двама речима – називом рода и описом

³⁵¹Alain Ray. *Essays on Terminology*. (Amsterdam : John Benjamins Publishing Co. 1995), 11-12

³⁵²Macmillan Dictionary. Nomenclature, <http://www.macmillandictionary.com/dictionary/british/nomenclature> (приступљено 17.11.2015)

³⁵³Милан Вујакија. *Лексикон страних речи и израза*. (Београд: Просвета, 1954), 945

³⁵⁴Vladimir Anić i Ivo Goldstein. *Rječnik stranih reči i izraza*. (Zagreb: Novi Liber, 2004)

врсте).³⁵⁵ Рад на номенклатури, не само у области биологије, већ и других наука и њихових дисциплина допринео је настанку терминологије као посебне области.

Како 20. век доноси развој техничких и инжењерских наука тако научници из тих области укључују у рад на попису појмова из својих научних области, а како техника, технологија и инжењерство изузетно брзо напредују јавља се потреба не само за именовањем нових појмова и концепата, већ и за терминологијом као посебном научном области.

Иако би нам данас било логично да су се за област терминологије најпре заинтересовали лингвисти, у пракси то није био случај. Чак ни научници из других хуманистичких и друштвених наука нису показивали интересовање за рад у овом подручју. Терминологија у данашњем смислу почиње да се развија тек тридесетих година 20. века са појавом радова Аустријанца Еугена Вистера (Eugen Wüster 1898-1977). Иако је Вистер долазио из поља техничких наука, као инжењер електротехнике своју докторску дисертацију из 1931. године под називом *Internationale Sprachnormung in der Technik, besonders in der Electrotechnik*³⁵⁶ је посветио терминологији. Вистерова дисертација ће постати фундаментални рад из ове области јер је у њој постављен низ принципа за рад са терминима и фундирана методологија за обраду терминолошких података.

Током свог професионалног живота, Вистер се, паралелно са професорским радом у области техничких наука на Универзитету у Бечу, бавио истраживањима из терминологије и лексикографије, израдом специјализованих речника, проучавањем есперанта, реформом ортографије немачког језика, као и стандардизацијом из области библиотекарства и информатике, библиографије и децималне класификације.³⁵⁷ О његовој огромној научној активности сведочи и библиографија са преко 500 самосталних радова,³⁵⁸ као и учешће у бројним међународним организацијама и пројектима који су се бавили терминологијом. Треба истаћи да је Технички одбор за

³⁵⁵ Margaret J. Anderson. *Carl Linnaeus: Father of Classification*. (Berkeley Heights, NJ: Enslow Publishers, 1997), 6-8

³⁵⁶ *Међународна стандардизација језика технике, посебно електротехнике* (превод аутора)

³⁵⁷ Erhard Oeser (Hrsg.). *Eugen Wüster (1898–1977): Leben und Werk, ein österreichischer Pionier der Informationsgesellschaft*. (Vienna : TermNet, 1998)

³⁵⁸ Ángela Campo. „The Reception of Eugen Wüster’s Work and the Development of Terminology.” (Dissertation. Université de Montréal, Faculté des études supérieures, 2012):1-2, https://papyrus.bib.umontreal.ca/xmlui/bitstream/handle/1866/9198/Campo_Angela_2012_these.pdf (приступљено 17.11.2015)

стандардизацију терминологије при ISO основан захваљујући Вистеровим истраживањима, као и да је Вистер иницирао настанак Међународног информационог центра за терминологију (International Information Centre for Terminology – Infoterm³⁵⁹) при Унеску.

6.1.2 Дефиниција појма

Појам *терминологија* потиче од немачке кованице *Terminologie* коју је у виду придева *terminologisch* 1788. године први пут употребио Кристијан Готфрид Шиц (Christian Gottfried Schütz 1747-1832), професор на универзитетима у Халеу и Јени, а која потиче од латинске речи *terminus* и грчке речи *λογία*. Према Oxford English Dictionary,³⁶⁰ прва употреба појма у енглеском језику датира из 1801. године.

Терминологија се у различитим међународним речницима и лексиконима дефинише на сличан начин, почев од тога да су то „речи или изрази који се користе у одређеном послу, науци или професији”,³⁶¹ преко тога да је то „1) скуп термина који припадају науци, технологији или било којој омеђеној интелектуалној области; 2) наука која се бави проучавањем термина и њиховом употребом”,³⁶² „номенклатура одређене области или науке”,³⁶³ „систематично проучавање термина... чија је намена да именују класе објеката и концепте”,³⁶⁴ до тога да је у питању „скуп речи, изрази или симбола који представљају концепте или предмете који се користе у одређеној области истраживања, науке или активности и чије је значење [...] јасно дефинисано, а некада су и публиковани у облику речника или лексикона.”³⁶⁵

Аутори домаћих лексикографских издања су веома рано појам *терминологија* препознали као битан и уврстили га у своје пописе. Већ 1937. године терминологија се дефинише као „систематизација или утврђивање и дефинисање значења гл. [главне]

³⁵⁹International Information Centre for Terminology – Infoterm, доступно на: <http://www.infoterm.info/>

³⁶⁰Oxford English Dictionary, доступно на: <http://www.oed.com/>

³⁶¹Macmillan Dictionary. Terminology, <http://www.macmillandictionary.com/dictionary/british/terminology> (приступљено 18.11.2015)

³⁶²The Online Dictionary of Language Terminology (ODLT). Terminology, <http://odlt.org/> (приступљено 18.11.2015)

³⁶³Webster's Third New International Dictionary of the English Language. Terminology. Unabridged ed. (Springfield, Mass., G. & C. Merriam Co, 1966)

³⁶⁴Le Petit Robert. Terminologie. (Paris: Robert, 1978)

³⁶⁵Joan M. Reitz. ODLIS – Online Dictionary for Library and Information Science. Terminology, <http://www.abc-clio.com/ODLIS/searchODLIS.aspx> (приступљено 18.11.2015)

речи, имена и појмова неке научне или култ. [културне] области.”³⁶⁶ У другим публикацијама срећемо дефиниције да јетерминологија, „вештачки говор, скуп вештачки направљених (стручних) израза (*термина*) неке научне области”,³⁶⁷ „скуп стручних израза једне научне области”,³⁶⁸ „систем назива, термина у некој научној области, номенклатура”,³⁶⁹ те „систематизован скуп, попис термина у некој стручној области”.³⁷⁰

Све ове дефиниције имају заједнички садржалац, а то је да је терминологија скуп термина, речи или израза из одређене научне области. Чињеница да је истакнуто да термин припада одређеној научној области указује на то да се не ради о речима из свакодневног говора или општег речника. Ипак, да ли и у коликој мери термини из једне научне области постају део свакодневног говора или општег речника српског језика анализираћемо у одељку 6.3.2 на примеру термина из области библиотекарства и информатике.

Пошто неке од дефиниција указују на то да терминологија може означавати и науку, у наредном одељку ћемо представити развој терминологије као науке.

6.1.3 Терминологија као наука

Развој терминологије као модерне научне дисциплине прошао је кроз неколико фаза. Следећи периодизацију у раду Пјера Ожеа (Pierre Auger)³⁷¹ коју истичеи Тереза Кабре (Teresa M. Cabré)³⁷², еволуцију модерне терминологије поделићемо у четири фазе:

- 1) постанак (1930-1960) – период настанка првих теоријских радова из области терминологије и стварање метода за систематско формирање термина;
- 2) структурирање области (1960-1975) – у овом периоду изузетно је значајна појава рачунара и развој техничке документације, те првих банки података што

³⁶⁶Свезнање: општи енциклопедиски лексикон. Терминологија (Београд: Народно дело, 1937)

³⁶⁷Милан Вујаклија. *Лексикон страних речи и израза*. (Београд: Просвета, 1954), 945

³⁶⁸Мала енциклопедија Просвете: општа енциклопедија, Т. 2. Терминологија (Београд: Просвета, 1959)

³⁶⁹Речник српскога језика. Терминологија. (Нови сад: Матица српска, 2007)

³⁷⁰Милица Радовић-Тешић. *Граматицки и лингвистички појмовник*. (Београд: Учитељски факултет, 2011), 144

³⁷¹Pierre Auger. „La terminologie au Quebec et dans le monde, de la naissance à la maturité.” In: *Actes du sixième colloque OLF-STQ de terminologie. L'ère nouvelle de la terminologie*, pp. 27-59. (Quebec: Gouvernement du Québec, 1988)

³⁷²Teresa M. Cabré. *Terminology: Theory, Methods, and Applications*. (Amsterdam : John Benjamins Publishing Co, 1999), 5

је све довело до координације у области обраде и стандардизације терминологије на међународном нивоу, као и израде првих тезауруса уз помоћ рачунара;

- 3) успон (1975-1985) – у ових десет година запажа се развој пројеката који се баве планирањем и модернизацијом језика који су засновани на знањима из терминологије, као и развој персоналних рачунара који доносе велики напредак у смислу обраде терминолошких података;
- 4) ширење (од 1985) – од средине осамдесетих година па надаље од велике је важности развој рачунарских наука и њихова примена у области терминологије јер су језички ресурси и алати, подржани рачунарским технологијама, допринели да терминологија значајно напредује као наука, да се олакша рад на пословима стандардизације термина из разних научних области, те да се координира развој терминологије у оквиру различитих међународних пројеката, организација и иницијатива;

Имајући у виду да је поменути рад Пјера Ожеа настао 1988, а публикација Терезе Кабре 1999, ову периодизацију потребно је допунити још једном фазом:

- 5) ширење на интернету (од 2000) – у овој фази еволуције терминологије доминантан је развој апликација које се ослањају на онлајн терминолошке речнике, базе података термина и тезаурусе из различитих научних области подржаних савременим рачунарским алатима и једнојезичним и вишејезичним лексичким ресурсима.

Развој терминологије као посебне научне дисциплине, дакле, директно зависи од развоја технологије и масовних комуникација, али и потребе за међусобном комуникацијом припадника научне заједнице.

Терминологија је примењена наука јер су се упоредо развијали и њени теоријски принципи и практична примена. Гофен³⁷³ сматра да терминологија у пракси еволуира кроз израду речника и лексикона који су њен директни резултат. Према Кабреовој (Cabré 1999), терминологија је област која има интердисциплинарни карактер и у тесној је вези са лингвистиком (примењена лингвистика, језички систем, лексикологија, лексикографија), когнитивним наукама (анализа термина, креирање концепата, односи између концепата, предметна класификација), комуникационим наукама (специјализована комуникација, превођење, планирање језика),

³⁷³Goffin, Roger. 1985. „La science terminologique.” *Terminologie et traduction*, 2(1985):9-29

документацијом, рачунарским наукама и изградњи система заснованих на управљању знањем. На интердисциплинарност терминологије указивао је још и Вистер који је истакао да је терминологија саставни део лингвистике (језичке јединице – проучавање термина), логике и онтологије (когнитивни елементи – везе између појмова и концепата), информационих наука (специјализована комуникација – аутоматизовани процеси процесирања текста, машинског превођења и сл.), а нарочито многих области рачунарских наука.³⁷⁴

Терминологија је, дакле, посебна научна дисциплина, примењена наука интердисциплинарног карактера на коју нарочито има утицај развој технологије и рачунарских наука.

Однос терминологије и рачунарских наука

Развој рачунарских технологија довео је до тога да је поједине послове много једноставније обављати него раније и не само то – раст и развој појединих научних дисциплина у директној је вези са иновативним процесима у креирању рачунарских програма и апликација. Такав је случај и са терминологијом која користи добра искуства рачунарски подржаних технологија у подручју обраде природног језика.

Данас је незамисливо вршити квалитетно истраживање већег обима из домена терминологије без примене знања из рачунарских наука. Историјски гледано, у свакој фази развоја рачунарских апликација које су се бавиле језиком и текстом, те апликације су нашле примену у лингвистици, а често и у проучавању терминологије.

У првој фази, језички подаци и термини су се користили у датом облику, без даље лексичке обраде и анализе. Програми који су се бавили провером ортографије (енг. *spelling checkers*) или растављањем речи на крају реда, односно хифенацијом представљају прве кораке примене рачунарских технологија како у обради текста, тако и у подручју терминологије. Друга фаза доноси рачунарске алате који се примењују на језик – израда електронских речника, рачунарски подржани системи за превођење и корекцију текста и сл. У трећој фази запажамо развој програмских апликација у смеру обраде и анализе података, тако да се у подручју терминологије све више користе анализатори, лематизери, класификатори и програми за обраду

³⁷⁴ Eugen Wüster. „L'étude scientifique générale de la terminologie, zone frontalière entre de la linguistique, la logique, l'ontologie, l'informatique et les science des choses.” In: H. Rondeau & G. Felber (eds.) *Textes choisis de terminologie. Vol. 1, Fondements théorétiques de la terminologie.* (Ville de Québec: Université Laval, GIRSTERM– Groupe interdisciplinaire de recherche scientifique et appliquée en terminologie, 1981):55-113.

статистичких података. Четврта фаза се односи на примену сложенијих рачунарских програма и алата који обухватају послове аутоматске идентификације термина и њихову екстракцију из текста, машинско превођење, системе за рачунарски подржано учење, аутоматско индексирање, односно на пресудан утицај напретка технологије на развој рачунарске лингвистике.³⁷⁵

Данас ти системи и апликације још увек нису довољно успешни да би у потпуности могли да замене човека јер и даље постоји потреба за коначном ручном провером пре закључења неког од побројаних процеса. Ипак, све је мање потребно интервенисати јер тимови стручњака разних профила раде на томе да апликације и алате доведу до апсолутне успешности. Такође, футуристичка визија Пјера Ожеа из 1989. године³⁷⁶ да ће научници који се баве терминологијом приступати великим базама података и преузимати термине који се налазе у корпусима, да ће процес бити потпуно аутоматизован и да неће бити потребе за ручним интервенцијама, а да ће се семантички дескриптори додељивати аутоматски и тако опремљени термини користити за даље дефинисање појмова, данас је у значајној мери остварена у пракси.

У одељцима 6.3.1 и 6.3.2 показаћемо да је Ожеова визија у завидној мери остварена, а у неким сегментима и надограђена и у домену српског језика, на примеру аутоматске екстракције термина применом морфолошких речника на корпусу текстова из области библиотекарства и информатике.

6.1.4 Стандардизација у области терминологије

Било да радимо на изради једнојезичног или вишејезичног термилошког речника, термилошке базе података или само на изради краће листе термина, да би тај рад био квалитетан, стручно утемељен и намењен потенцијалној размени са другим корисницима или системима, неопходно је да поштујемо стандарде. Као и у другим областима, рад у области терминологије подлеже одређеној стандардизацији прописаној на међународном, домаћем и стручном нивоу.

У погледу терминологије из области библиотекарства и информатике размотрићемо стандардизацију назначену у Међународној организацији за

³⁷⁵Teresa M. Cabré. *Terminology: Theory, Methods, and Applications*. (Amsterdam : John Benjamins Publishing Co, 1999), 160-161

³⁷⁶Pierre Auger. „La terminotique et les industries de la langue.” *Meta: journal des traducteurs*, 34,3(1989):450-456, <https://www.erudit.org/revue/meta/1989/v34/n3/001922ar.pdf> (приступљено 27.11.2015)

стандардизацију (ISO), Међународној федерацији библиотекарских удружења (IFLA) и Српском стандарду (СРПС, раније Југословенском стандарду ЈУС).

ISO стандардизација

Технички одбор за стандардизацију терминологије при ISO задужен је за доношење и вођење стандарда везаних за терминологију. Опис основног стандарда везаног за терминологију ISO 704:2009³⁷⁷ каже да овај стандард утврђује основне принципе и методе за припрему и прикупљање терминологије унутар и ван оквира стандардизације, и описује везе између објеката, концепата и њихових терминолошких репрезентација. Он такође успоставља опште принципе који регулишу формирање ознака и формулацију дефиниција, а ови принципи се односе на терминолошки рад у научним, технолошким, индустријским, административним и другим областима знања.

Поред овог основног стандарда којим се прописује рад на терминологији, каталог ISO стандарда из области терминологије³⁷⁸ доноси још четрдесетак стандарда који се, између осталог, баве усаглашавањем концепата и термина, израдом речника, представљањем одредница у речницима, алфабетским редоследом одредница у вишејезичким пописима термина, библиографским референцама и идентификаторима извора везаним за терминологију, рачунарским апликацијама везаним за терминологију, управљањем лексичким ресурсима (низ стандарда и подстандарда који се тичу оквира за морфосинтаксичке анотације, лингвистичке анотације, лексичко означавање, вишејезичне информације, семантичке анотације итд.), системима за управљање терминологијом, знањем и садржајем, сервисима за превођење и сл.

IFLA

Организација IFLA, као кровна институција у области библиотекарства и информатике, састоји се од 44 секције са специфичним дужностима и пољима интересовања. Усаглашавање терминологије, рад на терминолошким речницима, израда индекса и тезауруса за потребу предметне класификације и њихову примену у каталозима, библиографијама и разним врстама докумената укључујући електронске документе, задатак је Секције за класификацију и индексирање при IFLA-и.

³⁷⁷ ISO. Terminology work – Principles and methods, 2009. Ref. ISO 704:2009, http://www.iso.org/iso/catalogue_detail.htm?csnumber=38109 (приступљено 27.11.2015)

³⁷⁸ ISO. Standards catalogue– 01.020: Terminology (principles and coordination), including terminography, http://www.iso.org/iso/catalogue/catalogue_ics/catalogue_ics_browse.htm?ICS1=1&ICS2=20 (приступљено 27.11.2015)

Оперативни поступци при усаглашавању и стандардизацији терминологије воде се кроз различите радне групе у оквиру ове Секције у којима учествују библиотечки стручњаци из домена класификације и индексирања из целог света.

Водећи пројекти Секције везани за терминологију јесу Упутства за израду вишејезичких тезауруса (*Guidelines of Multilingual Thesauri*³⁷⁹) и Вишејезични речник каталожних термина и концепата (*Multilingual Dictionary of Cataloguing Terms and Concepts – MulDiCat*³⁸⁰) који садржи термине, концепте и дефиниције из области каталогизације на енглеском као изворном језику и њихове преводе на 25 језика, а међу њима и на српски језик.

Приликом израде различитих препорука, вишејезичних тезауруса и вишејезичних речника, чланови радних група при IFLA-и поштују стандарде које прописује ISO и упутства и прописе које је донео Унеско. На тај начин остварена је уједначеност у погледу креирања терминолошких одредница и њихових дефиниција и омогућена размена терминолошких информација са другим базама података и системима.

СРПС

Још у време постојања Југославије, ЈУС стандард је био усаглашен са међународном стандардизацијом и у погледу терминологије. Године 1979. објављени су стандарди *Терминологија – Појмови и називи: општи принципи* и *Терминологија – Системи појмова и њихово приказивање, Информације и документација – Смернице за састављање и даљи развој једнојезичких тезауруса*, а 1994. године и *Информације и документација – Упутства за успостављање и развој једнојезичких тезауруса* и *Информације и документација – Упутства за успостављање и развој вишејезичких тезауруса*, усаглашени са ISO стандардом. Ови стандарди су касније измењени и надограђени, па се у изворном облику више не користе, али је свакако јако важна чињеница да је још седамдесетих година постојала потреба за развојем стандарда у области терминологије.

У препорукама за даљи развој информатике у нашој земљи 1997. године истиче се значај стандардизације везане за српски језик и садржај на српском језику, па је дефинисано да државни органи треба да подрже развој посебних софтверских алата за

³⁷⁹ IFLA Classification and Indexing Section. *Guidelines of Multilingual Thesauri*, <http://archive.ifla.org/VII/s29/pubs/Profrep115.pdf> (приступљено 19.1.2016)

³⁸⁰ IFLA Classification and Indexing Section. *Multilingual Dictionary of Cataloguing Terms and Concepts (MulDiCat)*, <http://www.ifla.org/publications/multilingual-dictionary-of-cataloguing-terms-and-concepts-muldicat?og=70> (приступљено 19.1.2016)

српски језик, а пре свега да стандардизују и уреде основу за јединствену примену ових алата за целокупан југословенски културни садржај на различитим рачунским платформама.³⁸¹

Југословенски стандард је после распада бивше државе замењен српским стандардом (СРПС). Српске стандарде доноси Институт за стандардизацију Србије,³⁸² а стандарде израђују комисије намењене специфичним питањима. Комисија за терминологију (ознака у СРПС је А037) којом руководи проф. др Душко Витас задужена је за израду и усаглашавање стандарда из ове области. Комисија за терминологију је од успостављања СРПС стандарда донела следеће стандарде³⁸³ из наведене области:

- Услуге превођења – Захтеви за услуге (СРПС ЕН 15038:2008) – стандард се односи на „обезбеђење квалитета, следљивост, оцењивање усаглашености и сертификацију” преводилачких услуга;
- Терминолошки рад – Вокабулар – Део 1: Теорија и примена (СРПС ISO 1087-1:2003) – стандард доноси системски опис појмова у области терминологије;
- Терминолошки рад – Вокабулар – Део 2: Примене рачунара (СРПС ISO 1087-2:2005) – стандардом су дефинисани термини за обраду језика и информација везани за терминолошки рад;
- Приказивање/представљање одредница у речницима – Захтеви, препоруке и информације (СРПС ISO 1951:2009) – стандардом се спецификује генеричка структура једнојезичких, вишејезичких, општих и специјализованих речника, без обзира на медиј на коме се појављују, а у циљу лакшег проналажења, екстракције, размене и дисеминације лексикографских информација;
- Терминолошки производи и услуге – Преглед и упутство (СРПС ISO 22128:2011) – стандардом се истиче терминолошка пракса коју треба поштовати у пружању терминолошких услуга и производа из ове области;
- Оцењивање и утврђивање референтних вредности (бенчмаркинг) терминолошких ресурса – Општи појмови, принципи и захтеви (СРПС ISO 23185:2010) – стандардом се описују основни појмови који се тичу ефективне

³⁸¹ *Стратегија даљег развоја информатике у Савезној Републици Југославији (пројекат за Савезну владу)*. (Београд: Савезно министарство за развој, науку и животну средину, 1997), 38

³⁸² Институт за стандардизацију Србије, доступно на: <http://www.iss.rs/rs>

³⁸³ Институт за стандардизацију Србије. *Стандарди комисије А037*, http://www.iss.rs/rs/standard/index.php?national_committee_id=552 (приступљено 19.1.2016)

употребе термиолошких података, са освртом на потребе корисника тих података;

- Термиолошка политика – Развој и имплементација (СРПС ISO 29383:2012) – стандард је намењен креаторима политике и стратегије у области терминологије било да долазе из владиног сектора или непрофитних или комерцијалних организација;
- Термиолошки рад – Принципи и методе (СРПС ISO 704:2005) – стандард дефинише основне принципе за формулисање дефиниција термина и веза између предмета, појмова и њиховог представљања употребом терминологије;
- Практичне смернице за социотерминологију (СРПС ISO/TR 22134:2010) – технички извештај којим се дају смернице за социотермиолошке принципе, методе и вокабуларе.

Већ из ознака ових стандарда видимо да су настали на основу ISO стандардизације и да су усаглашене са конкретним ISO стандардом. Бројчана ознака СРПС стандарда идентична је оној у ISO стандарду, разлика се јавља само у години када је стандард донет. Такође, колико је стандардизација значајна за област терминологије јасно је и из овог пописа и описа донетих СРПС стандарда где видимо које је све аспекте терминологије потребно стандардизовати.

Развој рачунарских апликација и коришћење могућности које оне нуде у домену терминологије намеће потребу за доношењем нових стандарда. ISO стандардизација се већ увелико бави овим питањима и доноси стандарде о примени рачунарских апликација у домену терминологије, ти стандарди треба да се преведу и на наш језик и прилагоде, а домаћи стандарди треба да прате препоруке и добру праксу међународне стандардизације.

6.2 Терминологија из области библиотекарства и информатике

Специфичност терминологије

Као и у многим другим научним областима које су везане за развој технике и технологије, а нарочито за напредак рачунарских технологија, терминологија у области библиотекарства и информатике свакодневно трпи промене које се огледају у употреби нових техничких термина, акронима и скраћеница.³⁸⁴ Развој терминологије у

³⁸⁴ Mary Mortimer. *Library Speak: a Glossary of Terms in Librarianship and Information Management*. (Friendswood, USA: Total Recall Publications, 2007), IV

овој области прешао је дуг пут пратећи промене у делатности од традиционалне до дигиталне библиотеке окрећући се све више рачунарским наукама. Тако можемо запазити да се у последњих пар деценија фокус научних проучавања померио „од индексирања, каталогизације, класификације, библиографије, архивирања ка информационим технологијама, проналажењу информација, информационој писмености, дигитализацији, очувању дигиталног наслеђа, информационим сервисима, менаџменту, маркетингу, издаваштву, библиометрији, наукометрији итд.”³⁸⁵

У складу са наведеним је и анализа Љиљане Ковачевић која истиче да класично библиотекарство у традиционалној библиотеци „обухвата корпус који, поред [...] поља: набавка, каталогизација, класификација, циркулација докумената и информација, обухвата и следеће дисциплине: архивистику, библиографију, графику, дипломатику, издаваштво, историју књиге, књижарство, књиговештво, фотографију, штампарство” док је „информатичко доба донело, а библиотечка терминологија инкорпорирала термине који спадају у: електронске изворе, електронску обраду података, информатичке производе и услуге, информационе системе, масовне комуникације, мултимедије, рачунарску опрему, софтвер, телекомуникације.”³⁸⁶

Битан аспект развоја и промене терминологије у овој области огледа се у преклапању ове научне области са другим сродним дисциплинама – од архивистике, палеографије, документалистике, музеологије до теорије информација, проналажења информација, програмирања, аутоматске обраде текста, рачунарске лингвистике. Илустрације ради, у области библиотекарства и информатике као термини равноправно стоје „каталожки листић”, „папирус”, „бушена картица”, „корисник”, „електронски каталог”, „културна баштина”, „дигитални објекат”, „библиографски опис”, „метаподатак”, „управљање знањем”, „штампарска преса”, „електронско издаваштво”, „клинасто писмо”, „оптичко препознавање карактера” и др.

Увидом у ову кратку листу термина можемо уочити да је неке од њих лако могуће пронаћи и у најстаријим терминолошким речницима из поменуте области, али и да

³⁸⁵ Aleksandra Trtovac and Jelena Andonovski. „Enrichment of Morphological Dictionary of MWUs.” In: Pavlović-Lažetić, Gordana et al. (eds.), *Natural Language Processing for Serbian*, pp. 27-40. (Belgrade: University of Belgrade, Faculty of Mathematics, 2014):28-29.

³⁸⁶ Ljiljana Kovačević. „Srpska bibliotekarska terminologija u međunarodnom projektu: višezjezički rečnik bibliotekarske terminologije.” *Glasnik Narodne biblioteke Srbije* 1(2000), https://www.nb.rs/view_file.php?file_id=512 (приступљено 30.11.2015)

савремени тренутак и развој терминологије захтева допуну термилошког речника што је и један од циљева овог истраживања.

Важност стандардизације терминологије из библиотекарства на српском језику уочена је у нашој стручној јавности још шездесетих година 20. века, када су наши стручњаци, подстакнути сличним подухватима у свету, започели рад на изради термилошких речника из ове области. Изради првих речника претходиле су бројне дискусије и прилози објављени у часопису „Библиотекар.”

Први такав прилог је из 1961. године аутора Смиље Мишић, управнице Библиотеке САНУ, у коме се говори о неопходности термилошког речника из библиотекарства који би био енциклопедијског карактера и методологији његове израде („прикупити грађу из књига, часописа, али и из праксе, из живота”), те о нужности тимског рада при реализацији таквог посла („коначну редакцију морао би поред библиотекара дати лингвист”).³⁸⁷ Своја теоријска истраживања преточена у практични рад, Смиља Мишић крунисала је у *Грађи за библиотекарски термилошки речник српскохрватски-француско-немачко-руски*, али је та грађа, иако од библиотечких стручњака оцењена као изузетно квалитетна, до данас остала необјављена.

Смиља Мишић иницирала је и дискусију о раду на стандардизацији терминологије на српскохрватском језику и изради термилошког речника и предложила да дискусија буде објављена у часопису „Библиотекар” што је и учињено током 1964. и 1965. године.³⁸⁸ Дискусије су биле обимне, веома значајне, пуно су обећавале, али нису резултирале конкретним решењима или израдом речника.

Да бисмо показали важност праћења и стандардизације терминологије из области библиотекарства и информатике, кроз историјски осврт ћемо указати на најзначајније речнике, лексиконе и тезаурусе који су основа за развој терминологије у овој научној области, а који су издати и на српском, односно српскохрватском језику.

³⁸⁷Смиља Мишић. „Израда српскохрватског библиотекарског речника: термилошки проблеми.” *Библиотекар* 13,4(1961):227-241.

³⁸⁸Дискусија о библиотекарској терминологији. *Библиотекар* 16,5/6(1964):305-315; 17,1/2(1965):4-8; 17,3/4(1965):113-117; 17,5/6(1965):318-324.

Речници и лексикони*Vocabularium bibliothecarii*

Иако је на подручју вишејезичких термилошких речника из библиотекарства било успешних подухвата и раније^{389,390}, за библиотекарску терминологију на нашем језику први свеобухватнији и значајнији речник је *Vocabularium bibliothecarii* који је, под покровитељством Унеска, издат 1953. године. То издање чији је састављач Ентони Томпсон (Anthony Thompson), наслања се на материјал за тројезични француско-енглеско-немачки речник Анрија Леметра (Henri Lemaître, 1881-1946) и сарадника из 1939. године који је одобрио Међународни завод за интелектуалну сарадњу и послао на рецензију виђеним библиотечким стручњацима тог времена, али који није објављен због избијања Другог светског рата. У Леметровом речнику, грађа је била сложена абecedно, а изворни језик био је француски, док Томпсоново издање доноси одреднице сређене према струкама, на основу принципа Универзалне децималне класификације, а изворни језик постаје енглески. „У коликој мери се осећала потреба за једним оваквим речником говори и чињеница да је доштампано издање речника изашло 1954. године...”³⁹¹ Значај речника огледа се и у томе да је 1959. године тројезични речник преведен на италијански, турски и хебрејски, а да се друго издање са два нова језика – шпанским и руским, јавља 1962. године. Петојезични речник преведен је на српскохрватски 1965. године и то представља изузетно важан тренутак у развоју и стандардизацији терминологије из библиотекарства и информатике на нашем језику. У наредних десетак година, у

³⁸⁹ Нпр. 1) вишејезични речник на 7 језика– *Vocabulaire technique de l'éditeur en sept langues: français, deutsch, English, español, hollandsch, italiano, magyar.* (Bern: Congrès international des éditeurs, 1913); 2) вишејезички речник на 8 језика Axel Moth. *Glossary of Library Terms: English, Danish, Dutch, French, German, Italian, Spanish, Swedish.* (Boston: The Boston Book Co, 1915); 3) вишејезични речник на 20 језика – Barbara Cowles. *Bibliographers' glossary of foreign words and phrases; an alphabet of terms in bibliographical and booktrade use, compiled from twenty languages.* (New York, R.R. Bowker Co., 1935)

³⁹⁰ Први вишејезички термилошки речник у који је био укључен и српскохрватски језик био је *Dictionarium Bibliothecarii Practicum* који је у виду седам издања излазио од 1963. до 1977. године. У првом издању је било 20 језика, а изворни језик је био мађарски. Из Југославије су на речнику сарађивали стручњаци из Југословенског библиографског института у Београду и Националне и свеучилишне књижнице у Загребу, а српскохрватски и хрватскосрпски третиран су у речнику као посебни језици.

³⁹¹ Добра Бегенишић. „Стандардизација библиотечке терминологије уз примену нових технологија.” У: Вранеш, Александра и Марковић, Љиљана (ур.) *Србија између истока и запада : наука, образовање, култура, уметност : тематски зборник у 4 књиге. Књ. 4, Језици Балкана у компаративном и интердисциплинарном контексту.* (Београд : Филолошки факултет Универзитета, 2014)

Vocabularium bibliothecarii укључени су и холандски (1967), словачки (1970) и мађарски језик (1972).

Рјечник библиотекарских стручних израза

Као резултат превода петојезичног речника *Vocabularium bibliothecarii* године 1965. јавља се превод овог речника на хрватскосрпски језик под називом *Рјечник библиотекарских стручних израза*, а у издању „Школске књиге” из Загреба. За разлику од других издања Унесковог вишејезичког речника, овде је реч само о двојезичном речнику у коме је изворни језик енглески, а циљни језик хрватскосрпски, док је позитиван аспект тај што је, како је наведено у предговору,³⁹² заступљена сва терминологија из оригиналног речника. Иако је лоша страна речника то што није вишејезични, ипак је ово широко коришћен и опште познат приручник из области библиотекарства на простору Југославије.

Енциклопедијски лексикон библиотекарства

Први лексикон библиотекарства на нашем језику је *Енциклопедијски лексикон библиотекарства* аутора Косте Грубачића из 1964. године. У предговору овог лексикона је истакнуто да је намењен онима који се баве библиотечком, издавачком и књижевном делатношћу, али и библиофилима, те да се библиотекарство као наука у нашој земљи тек развија и да скоро да нема приручника из те области. Аутор у предговору разматра и проблематику терминологије, наводећи да „пошто се наука и техника стално развијају, стварају се нови појмови и изрази. Стога је требало да у лексикону буде заступљена савремена библиолошка терминологија. То представља посебну тешкоћу у погледу стварања израза на нашем језику. Стога су употребљени не само интернационални него понекад и посебни изрази (енглески, немачки и др.).”³⁹³

Одреднице су у овом лексикону сложене абецедно, а поштује се етимолошки принцип што значи да су страни називи и имена наведени у изворном етимолошком облику. Осим термина из библиотекарства и сродних области, лексикон садржи и један број књижевних термина. Мада поседује извесне недостатке јер доноси податке о личностима, листовима, часописима који су афирмисани у сфери библиотекарства

³⁹²Antony Thompson. *Rječnik bibliotekarskih stručnih izraza: englesko-hrvatskosrpski*. (Zagreb: Školska knjiga, 1965)

³⁹³Kosta Grubačić. *Enciklopedijski leksikon bibliotekarstva*. (Sarajevo: Zavod za izdavanje udžbenika, 1964)

до краја 19. века, али не и о савременицима и савременим гласилима, ово је изузетно значајан приручник јер је први такве врсте на нашем језику.

Терминолошки речник из области научних информација

Овај вишејезични терминолошки речник из области научних информација садржи 1281 појам сређених у алфабетском редоследу. Верзија у коју је укључен и српскохрватски језик објављена је 1969. Године,³⁹⁴ захваљујући подршци Савезног савета за координацију научних делатности СФРЈ и садржи појмове на руском језику и њихове еквиваленте на српскохрватском, македонском и словеначком језику.³⁹⁵

Библиотекарски лексикон

Терминолошки рад у области библиотекарства и информатике на српском језику није имао значајније резултате све до 1984. године када је објављен *Библиотекарски лексикон* који су израдили Владимир Јокановић, Фахрудин Календер, Емил Поповић и Милош Стојановић.³⁹⁶ Иако је овај лексикон имао сасвим солидног домаћег претходника у виду *Енциклопедијског лексикона библиотекарства* и међународно издање *Рјечника библиотекарских стручних израза*, аутори се у попису одредница ограничавају само на област библиотекарства без појмова из сродних и рубних области, одузимајући тако свом приручнику интердисциплинарни карактер, а самим тим умањујући и значај који приручник ове врсте треба да има. У складу са тим је и коментар Добриле Бегенишић о овом приручнику да „сужавање терминологије на уско стручну [...] сужава предмет проучавања дисциплине која је у том периоду ионако била недовољно развијена.³⁹⁷ ”Међутим, „како су сва ова прегнућа била углавном резултат индивидуалног рада, то су и извесни пропусти разумљиви и

³⁹⁴ А. И. Михайлов (ред. кол.). *Терминологический словарь по научной информации : (на русском, сербохорватском, словенском и македонском языках) = Terminološki rečnik iz oblasti naučnih informacija : (na ruskom, srpskohrvatskom, slovenačkom i makedonskom jeziku) = Terminološki slovar iz područja znanstvene informacije : (v ruskem, srbohrvatskem, slovenskem i macedonskem jeziku) = Терминолошки речник за научна информација : (на руски, српско-хрватски, словеначки и македонски јазик)* (Београд: Институт за научно-техничку документацију и информације, 1969)

³⁹⁵ У изради и опису термина на српскохрватском језику учествовала је, поред бројних коаутора, др Десанка Стаматовић, знаменити професор и један од оснивача Катедре за библиотекарство и информатику Филолошког факултета.

³⁹⁶ Vladimir Jokanović, Fahrudin Kalender, Emil Popović i Miloš Stojanović. *Bibliotekarski leksikon*. (Београд: Nolit, 1984)

³⁹⁷ Добрила Бегенишић. „Стандардизација библиотечке терминологије уз примену нових технологија.” У: Вранеш, Александра и Марковић, Љиљана (ур.) *Србија између истока и запада : наука, образовање, култура, уметност : тематски зборник у 4 књиге. Књ. 4, Језици Балкана у компаративном и интердисциплинарном контексту*. (Београд : Филолошки факултет Универзитета, 2014)

прихватљиви, занемарљиви пред великим успехом отпочињања посла дефинисања појмова који води стандардизацији терминологије, као једном од основних предуслова утемељења једне науке...”³⁹⁸

Мала југословенска енциклопедија библиотекарства

Аутори *Библиотекарског лексикона*, са изузетком Фахрудина Календера, наставили су рад на терминологији из области библиотекарства и 1988. године објавили *Малу југословенску енциклопедију библиотекарства*.³⁹⁹ Поново је проблем са приручником тај што се ограничава само на област библиотекарства, не сагледавајући ову научну област у ширем контексту. Предност ове енциклопедије је у томе што се односи на библиотекарство у свим републикама Југославије – одреднице за библиотеке су места у којима се библиотеке налазе, сви прописи који се односе на послератно југословенско библиотекарство дати су хронолошки, а дат је и детаљан преглед доктора и магистара библиотекарства, као и аутора стручних монографија. Из наведених разлога, приручник је добар покушај систематизовања библиотекарства у Југославији, а тиме и оправдава наслов под којим је издат.

Библиотекарски термилошки речник: енглеско-српски, српско-енглески

Пошто је *Библиотекарски термилошки речник* ауторки Љиљане Ковачевић, Весне Ињац и Добриле Бегенишић из 2004. године један од лексичких ресурса који се користе за истраживања у оквиру ове докторске дисертације, он је већ детаљно описан у поглављу 4.4. На овом месту ћемо само додати да су у изради речника ауторке имале значајну консултативну помоћ стручњака из Института за српски језик, а да је пројекат финансирало тадашње Министарство финансија Републике Србије. С обзиром на мултидисциплинарни карактер речника, двојезичност, велики број добро систематизованих одредница, те постојање и штампане верзије и верзије на CD-у, као и редовно ажуриране онлајн базе, стручна јавност је у овом речнику препознала изузетно квалитетан приручник.

Од рукописа до библиотеке

Приручник *Од рукописа до библиотеке* ауторки Александре Вранеш и Љиљане Марковић је појмовник који доноси дефиниције термина из технике научног рада,

³⁹⁸ Александра Вранеш. „Термилошке двоумице у библиотекарству.” У: Младеновић, Добривоје (прир.), *Библиотекарство на крају века*, Књ. 4, стр. 65-68. (Београд: Библиотекарско друштво Србије, 1999):66

³⁹⁹ Vladimir Jokanović, Emil Popović i Miloš Stojanović. *Mala jugoslovenska enciklopedija bibliotekarstva*. (Beograd: Nova knjiga, 1988)

методологије истраживања, библиотекаштва, издаваштва и књижаства. Према Уводној речи,⁴⁰⁰ књига је настала у циљу што боље доступности информација из поменутих области, а садржи термине који, осим оних који се јављају у ранијим приручницима, укључују и бројне савремене појмове. О значају и популарности приручника међу стручњацима, али и студентима и истраживачима из наведених области довољно говори податак да је појмовник већ имао три издања (прво издање 2006, друго допуњено издање 2008, а треће допуњено и измењено издање 2013. године).

Тезауруси

У савременим системима за проналажење информација, тезауруси су незаобилазно средство у поступцима аутоматског индексирања. Тезауруси садрже терминологију која је усвојена и нормирана јер су, како је наведено у Смерницама стандарда ISO 2788:1986,⁴⁰¹ средство за терминолошку контролу. Увидом у базе података и каталог великих светских библиотека, утврдили смо да је израђено неколико свеобухватних тезауруса за област библиотекаштва и информатике (тезуруси на енглеском језику: Information Science Thesaurus,⁴⁰² Thesaurus of Information Science Terminology,⁴⁰³ ASIS Thesaurus of Information Science and Librarianship,⁴⁰⁴ ASIS&T Thesaurus of Information Science, Technology and Librarianship⁴⁰⁵; немачко-енглески тезаурус Semiotischen Thesaurus der Bibliothekwissenschaft⁴⁰⁶; шпанско-каталонско-енглеско-француски тезаурус Tesouro de Biblioteconomía y Documentación⁴⁰⁷), али се као основа за

⁴⁰⁰ Александра Вранеш и Љиљана Марковић. *Од рукописа до библиотеке*. (Београд: Филолошки факултет Универзитета, 2008)

⁴⁰¹ ISO 2788:1986 – Guidelines for the establishment and development of monolingual thesauri, http://www.iso.org/iso/catalogue_detail.htm?csnumber=7776 (приступљено 30.11.2015)

⁴⁰² Claire Schultz. *Information Science Thesaurus*. ([Philadelphia] Drexel Institute of Technology, 1966 [i.e. 1967])

⁴⁰³ Claire Schultz. *Thesaurus of Information Science Terminology*. (Washington, Communication Service Corp., 1968)

⁴⁰⁴ Jessica L. Milstread (ed.). *ASIS Thesaurus of Information Science and Librarianship*. (Medford, NJ : Published for the American Society for Information Science by Learned Information, 1994)

⁴⁰⁵ Alice Redmond-Neal and Marjorie M.K. Hlava (eds.). *ASIS&T Thesaurus of Information Science, Technology and Librarianship*. (Medford, N.J. : Published for the American Society for Information Science and Technology by Information Today, 2005)

⁴⁰⁶ Ben Kaden (hrsg.). *Semiotischen Thesaurus der Bibliothekwissenschaft*, 2003, доступно на: <http://www.ib.hu-berlin.de/~wumsta/infopub/textbook/definitions/thesauindex.html>)

⁴⁰⁷ *Tesouro de Biblioteconomía y Documentación*, 2002, доступно на: http://thes.cindoc.csic.es/index_BIBLIO_esp.php

терминологију из ове области често користе и специјализовани тезауруси – од предметних одредница Конгресне библиотеке,⁴⁰⁸ преко тезауруса чија је тема стара и ретка књига,⁴⁰⁹ египтологија,⁴¹⁰ књижевност⁴¹¹ до оних који се баве информационим технологијама или масовним комуникацијама.⁴¹²

Иако није терминолошки уско оријентисан на област библиотекарства и информатике, због историјског аспекта, вишејезичности и континуитета, битно је навести и Унесков енглеско-француско-шпанско-руски тезаурус из 1977. године који међу терминима из области образовања, културе, природних, друштвених и хуманистичких наука, комуникације и информисања, доноси и бројне термине из области библиотекарства и информатике, а данас има и онлајн верзију.⁴¹³

Ако размотримо употребу тезауруса у контексту библиотекарства и информатике и терминологије на српском језику, запазићемо да није израђен тезаурус за ову област који би се користио у процесима индексирања, проналажења информација или у изради предметних одредница у нашим библиотекама.

У ширем контексту вишејезичних тезауруса који садрже и српски језик,⁴¹⁴ поново нема специјализованог тезауруса из ове области. Терминологија из библиотекарства и информатике јавља се у оквиру вишејезичног тезауруса Европске уније – EuroVoc,⁴¹⁵ а српски је један од 26 језика који су садржани у овом речнику. Међутим, како попис термина из наведене области у тезаурусу EuroVoc није исцрпан, он може да послужи само као основа при изради терминолошког приручника.

⁴⁰⁸Library of Congress Subject Headings – LCSH, доступно на: <http://id.loc.gov/authorities/subjects.html>

⁴⁰⁹ American Library Association (ALA). *A Thesaurus for Use in Rare Book and Special Collections Cataloguing*. (Chicago: Association of College and Research Libraries, 1991)

⁴¹⁰Centre for Computer-aided Egyptological Research (CCER). *Multilingual Egyptian Thesaurus*, доступно на: <http://projetrosette.info/page.php?Id=290>

⁴¹¹ American Library Association (ALA). *Guidelines on Subject Access to Individual Works of Fiction, Drama (GSAFD)*, доступно на: <http://experimental.worldcat.org/gsafd/>

⁴¹²Jean Viet. *Thesaurus Mass Communication*. (Paris: UNESCO, 1975), доступно на: <http://unesdoc.unesco.org/images/0001/000150/015031MB.pdf>

⁴¹³ UNESCO Thesaurus, доступно на: <http://databases.unesco.org/thesaurus/>

⁴¹⁴Basel Register of Thesauri, Ontologies and Classifications – BARTOC, доступно на: <http://bartoc.org/en>

⁴¹⁵EuroVoc – Вишејезични речник Европске уније, доступно на: <http://eurovoc.europa.eu/drupal/?q=sr>

6.2.1 Допуна термилошког речника српског језика из библиотекарства и информатике

Стандарди за термилошке базе података

Код процеса изградње и одржавања термилошких база података изузетно је битна примена стандарда препоручених за ову област ради постизања термилошке доследности у једној или више термилошких база података, а и да би се омогућило да једна термилошка база комуницира са различитим софтвером у смислу преноса података, размене података и њихове поновне употребе. У контексту допуне електронског термилошког речника из библиотекарства и информатике на српском језику, потребно је описати који се стандарди користе у конкретном случају и на који начин су они примењени.

Међународни стандард TermBase eXchange (TBX⁴¹⁶) дефинисали су ISO⁴¹⁷ и Удружење за локализацију индустријских стандарда (Localization Industry Standards Association – LISA⁴¹⁸). Стандардом TBX дефинише се XML формат за размену термилошких података при чему термилошка база података представљена TBX-ом мора да буде у складу са Оквиром за означавање терминологије (Terminological Markup Framework – TMF⁴¹⁹), односно апстрактним моделом података који је такође дефинисао ISO.⁴²⁰ Стандард TMF нуди модел података за опис терминологије као и смернице о основним принципима за представљање података у термилошким базама. Овај оквир пружа модел и методе за опис специфичних језика за означавање терминологије (TMLs). TBX је једна од апликација TMF-а, а TBX спецификација конкретног TML-а описује који су различити типови термилошких података дозвољени за употребу и на којим нивоима термилошког уноса се они појављују.⁴²¹ XML документ у стандарду TBX користи формат MARTIF (Machine-Readable

⁴¹⁶Introduction to TermBase eXchange (TBX), доступно на: <http://www.tbxinfo.net/>

⁴¹⁷ISO. Systems to Manage Terminology, Knowledge and Content – TermBase eXchange (TBX), 2008. Ref. ISO 30042:2008

⁴¹⁸Localization Industry Standards Association – LISA, доступно на: <http://web.archive.org/web/20110101184308/http://www.lisa.org/>

⁴¹⁹Terminological Markup Framework – TMF, доступно на: <http://www.loria.fr/projets/TMF/>

⁴²⁰ISO. Computer Applications in Terminology – Terminological Markup Framework, 2003. Ref. ISO16642:2003

⁴²¹Ranka Stanković, Ivan Obradović and Miloš Utvić. „Developing Termbases under the TBX Standard.” In: Pavlović-Lažetić, Gordana et al. (eds.), *Natural Language Processing for Serbian*, pp. 12-26. (Belgrade: University of Belgrade, Faculty of Mathematics, 2014):18-19.

Terminology Interchange Format) за размену термилошких података, а који се дефинише ISO стандардом.⁴²² TBX се користи и у алатима за превођење заснованим на преводачким меморијама (енг. Translation memory или ТМ) заједно са ТМХ. Водећи алати за превођење као што су SDL Trados,⁴²³ МемоQ⁴²⁴ и сл. директно подржавају поменуте стандарде.

Термилошке базе података

Како наводи Алан К. Мелби (Alan K. Melby) TBX није неопходан при изградњи двојезичних речника са терминима у два ступца, али је довољно флексибилан да се користи и у таквим случајевима, али и при изради сложених термилошких база. Он такође истиче да ТМФ нуди веома флексибилан оквир за дефинисање структуре термилошке базе, али да је размена података јасно дефинисана.⁴²⁵ Флексибилност се огледа у томе да су од свих категорија обавезне само „термин” (енг. term) и „језик” (енг. language). Недостатак такве флексибилности је у томе што две термилошке базе истим категоријама могу именовати различите карактеристике, па то може донети сметњу у процесу размене термилошких података.

ТМФ је састављен од следећих делова:

- Одељак језика – сви термини на једном језику групишу се заједно; за ову врсту података користи се етикета <langSet>;
- Одељак термина – сваки термин, невезано за језик на ком је, има свој посебан одељак у коме су термини којима су додељени исти метаподаци јер припадају истој врсти; за ову врсту података користи се етикета <termEntry>, за сваки појединачни термин етикета <descript type=”relatedConcept”>; у оквиру етикете <termEntry> сваком концепту придружују се подређени и надређени концепт;
- Саставни делови термина – вишечлане речи су секвенције речи које имају свој род, број, флексију и прецизно одређено значење и као такве јављају се као посебни термини у термилошким базама; овај одељак служи томе да идентификује вишечлане речи и онемогући да се њихови саставни делови

⁴²²ISO. Computer Applications in Terminology – Machine-Readable Terminology Interchange Format (MARTIF) – Negotiated Interchange, 1999. Ref. ISO 12200:1999

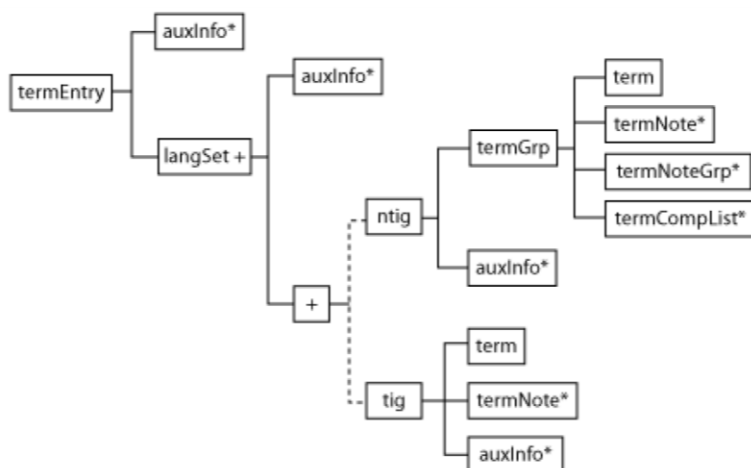
⁴²³SDL Trados, доступно на:<http://www.translationzone.com/products/sdl-trados-studio/>

⁴²⁴МемоQ, доступно на:<https://www.memoq.com/>

⁴²⁵Alan K. Melby. „Terminology in the Age of Multilingual Corpora.” *The Journal of Specialized Translation* 18(2012):7-29, http://www.ttt.org/tbx/art_melby-JoSTransNum18.pdf (приступљено 11.12.2015)

рашчлане; вишечлане речи уносе се као посебни елемент <termEntry>, а просте речи које чине вишечлану реч наводе се у оквиру елемента <termGrp>.

Структура једне термилошке одреднице са свим нивоима приказана је на слици 67.



Слика 67 – Нивои термилошке одреднице

Минималан захтев TMF-а је да термилошка база буде концептуално оријентисана, а не да одредница и сва њена значења буду груписани заједно. Концептуално оријентисана термилошка база подразумева хијерархијске односе – надређеност (хипероним), подређеност (хипоним) и једнакост (синоним), али и партитивне односе (односи целина-део – нпр. „глава-око”) и асоцијативне односе (однос термина и свих других језичких елемената са којим се он доводи у везу – нпр. „студент” у односу на „индекс”, „факултет”, „универзитет”, „апсолвент” и сл.).⁴²⁶ Концептуални односи представљају основу за семантички веб, а у термилошким базама управо стандард ТВХ омогућава ову врсту односа.

Термилошка база података из библиотекарства и информатике коју у овом истраживању претражујемо алатом *Библиша* садржи хијерархијске односе међу терминима, уз могућност да се приликом претраге укључе или искључе хипероними и хипоними. На примеру претраге то изгледа овако – ако је задати термин за претрагу *књига* и претрага укључује хиперониме и хипониме, онда као резултат претраге, осим термина *књига* на српском и *book* на енглеском језику, добијамо и следеће термине:

1. хипероними – *производ*, *публикација*, *том*, *волумен* и сл, као и еквиваленти на енглеском језику *product*, *publication*, *tome*, *volume* и сл.

⁴²⁶Ова врста веза наликује везама које се успостављају у семантичким мрежама типа wordnet

2. хипоними – *албум, бележница, блок за цртање, брошура, цртанка, језички водич* и сл, као и еквиваленти на енглеском језику *album, notebook, sketch block, sketch pad, sketchbook, booklet, brochure, drawing book, phrase book* и сл.

Оваква врста погодака могућа је због тога што ТВХ подразумева унос термина на нивоу подређених и надређених концепата који су дефинисани етикетама `<descript type="relatedConceptBroader">` и `<descript type="relatedConceptNarrower">`.

Показаћемо један пример уноса термина са хијерархијским везама у виду подређених и надређених концепата у ТВХ-у:

```
<termEntry id="A10001">
  <descrip type="subjectField">librarianship</descrip>
  <descrip type="relatedConceptBroader">publication</descrip>
  <descrip type="relatedConcept">library </descrip>
  <descrip type="relatedConceptNarrower">booklet</descrip>
  <langSet xml:lang="en">
    <ntig>
      <termGrp>
        <term>book</term>
        <termNote type="termType">entryTerm</termNote>
        <termNote type="grammaticalGender">masculine</termNote>
        <termNote type="partOfSpeech">noun</termNote>
      </termGrp>
      <descripGrp>
        <descrip type="definition">A set of printed sheets of paper that are held together inside a cover; a long written work that can be read on a computer; a set of sheets of paper that are inside a cover and that you can write information on.</descrip>
      </descripGrp>
    </ntig>
  </langSet>
  <langSet xml:lang="fr">
    <descrip type="subjectField">biblioth conomie</descrip>
    <descrip type="relatedConceptBroader">publication</descrip>
    <descrip type="relatedConcept">biblioth que</descrip>
    <descrip type="relatedConceptNarrower">brochure</descrip>
    <ntig>
      <termGrp>
        <term>livre </term>
        <termNote type="termType">entryTerm</termNote>
        <termNote type="grammaticalGender">masculine</termNote>
        <termNote type="partOfSpeech">nom</termNote>
      </termGrp>
      <descripGrp>
        <descrip type="definition">Un set de feuilles de papier imprim es qui sont maintenues ensemble   l'int rieur d'une couverture; un long travail  crit qui peut  tre lu sur un ordinateur; un set de feuilles de papier qui sont   l'int rieur d'une couverture et de quoi vous pouvez  crire des informations.</descrip>
      </descripGrp>
    </ntig>
  </langSet>
  <langSet xml:lang="sr">
    <descrip type="subjectField">bibliotekarstvo</descrip>
    <descrip type="relatedConceptBroader">publikacija</descrip>
    <descrip type="relatedConcept">biblioteka</descrip>
```

```

<descrip type="relatedConceptNarrower">brošura</descrip>
<ntig>
  <termGrp>
    <term>knjiga</term>
    <termNote type="partOfSpeech">noun</termNote>
    <termNote type="termType">entryTerm</termNote>
    <termNote type="grammaticalGender">feminine</termNote>
  </termGrp>
  <descripGrp>
    <descrip type="definition">Skup štampanih listova papira koji su povezani zajedno unutar korica; dugi pisani rad koji se može čitati na računaru; skup listova papira koji su unutar korica i o čijem sadržaju se može pisati. </descrip>
  </descripGrp>
</ntig>
</langSet>
</termEntry>

```

У наведеном примеру можемо уочити да је описан термин *књига* на енглеском (*book*), француском (*livre*) и српском језику, да термин припада предметној области *библиотекарство* (енг. *librarianship*, франц. *bibliothéconomie*), као и да су наведени његови надређени и подређени концепти (*публикација*, *брошура*; енг. *publication*, *booklet*; франц. *publication*, *brochure*). Структуром ТВХ-а дефинисано је да се најпре јављају концепти на енглеском језику јер се он сматра изворним језиком, па тек онда на осталим језицима термилошке вишејезичне базе података. Такође, ТВХ структура нуди два могућа сценарија за унос атрибута за енглески језик када је изворни језик (<langSet xml:lang="en">) – могућност да се атрибут за енглески језик јави иза уноса атрибута за предметну област (<descrip type="subjectField">), надређени (<descrip type="relatedConceptBroader">) и подређени концепт <descrip type="relatedConceptNarrower"> што је случај у наведеном примеру, као и да се овај атрибут нађе испред поменутих атрибута.

Термилошке базе података нуде две врсте структуре: 1) сви језици су равноправни или 2) један језик (најчешће енглески) је изворни језик, а остали језици су циљни језици. У домаћим вишејезичним термилошким базама података српски језик је најчешће изворни језик, а остали језици су циљни језици. Структура описана на слици 30 показује ситуацију када постоје изворни језик и циљни језици и таква јесте и структура термилошке базе Терми⁴²⁷ у оквиру које се врши унос термина потребних за ово истраживање.

⁴²⁷ TERMI, доступно на: <http://termi.rgf.bg.ac.rs/>

ТЕРМИ

Терминолошка вишејезична база, односно терминолошки речник Терми настао је и развија се у оквиру Темпус пројекта ВАЕКТЕЛ.⁴²⁸ Апликација садржи термине из различитих области (рударство, геологија, математика, информатика, библиотекарство и других), а састоји се од четири странице (прелиставање, претрага, ажурирање и профили). Странице за прелиставање и претрагу су јавно доступне, док је за приступ страницама за ажурирање и уређивање профила потребно имати кориснички налог са потребним привилегијама.

Апликација је замишљена као стабло са хијерархијском структуром термина, а језици на којима се врши унос и претрага су српски и енглески. Пошто ћемо у овом истраживању радити на страници за ажурирање термина, детаљније ћемо описати ову страницу апликације Терми. На овој страници могуће је извршити унос новог термина, односно допуну, измену или брисање постојећег термина. Како је у питању хијерархијска структура, сваки термин има свој надређени концепт („родитеља“), а најчешће постоје и подређени концепти („деца“). Унос новог термина захтева најпре одабир правог корена, односно „родитеља.“ Након одабира „родитеља“, апликација нуди маску за унос новог термина, која садржи термин на српском и енглеском језику, његове синониме и дефиниције термина на оба језика, као и предметну област којој термин припада. Маска за унос новог термина приказана је на слици 68.

Слика 68 – Маска за унос новог термина у базу Терми

Структуром апликације предвиђено је повезивање података о конкретном термину са библиографском референцом одакле је преузета дефиниција термина. Уколико конкретна библиографска јединица није раније унета у базу, потребно је унети

⁴²⁸ ВАЕКТЕЛ – Blending academic and entrepreneurial knowledge in technology enhanced learning, доступно на: <http://www.baektel.eu/>

податке о њој да би се повезала са описиваним термином. Маска за унос библиографских података приказана је на слици 69.

Bibliografija

Naslov	Bibliotekarski terminološki rečnik: englesko-srpski, srpsko-engleski
Skracenicica	
Autor	Ljiljana Kovačević, Vesna Injac, Dobrila Begenišić
Apstrakt	
Izdavač	Narodna biblioteka Srbije
Godina izdanja	2004
Broj strana	
Tip reference	Terminološki rečnik
Edicija	
Tom knjiga	
Seriya	
<input type="button" value="Snimi"/>	

Слика 69 – Маска за унос библиографских података о извору у бази Терми

Када је термин унет у базу и када су подаци о њему сачувани, потребно је да редактор верификује исправност унетог термина и све остале податке у вези са термином. Након верификације, могућ је извоз терминолошких података у виду Excel документа или документа у ТВХ формату (слика 70).

```
<termEntry id="c110742">
  <langSet xml:lang="sr">
    <ntig>
      <termGrp>
        <term>univerzitetska biblioteka</term>
        <termNote type="termType">entryTerm</termNote>
      </termGrp>
      <descripGrp><descrip type="definition">Univerzitetska biblioteka je matična biblioteka u mreži biblioteka univerziteta. Kao takva, ona ispunjava osnovne funkcije matičnosti i razvoja, odnosno pruža informaciono-referalne usluge korisnicima mreže i prikuplja i konzervira bibliotečki materijal iz mreže. Stara se, takođe, o razvoju sistema i o stručnom usavršavanju bibliotečkih radnika. Obavlja i druge poslove koji su važni za mrežu u celini.</descrip>
        </descripGrp>
    </ntig>
  </langSet>
</termEntry>
```

```

</langSet>
<langSet xml:lang="en">
  <ntig>
    <termGrp>
      <term>university library</term>
      <termNote type="termType">entryTerm</termNote>
    </termGrp>
    <descripGrp><descrip type="definition">A library or library system established, administered,
and funded by a university to meet the information, research, and curriculum needs of its
students, faculty, and staff. Some large universities maintain separate undergraduate and graduate
libraries. Large university libraries with comprehensive collections are considered research
libraries.</descrip>
    </descripGrp>
  </ntig>
</langSet>
</termEntry>

```

Слика 70 – Извоз термилошких података из базе Терми у ТВХ формату

6.2.2 Примена морфолошких речника на корпусу текстова

Екстракција вишечланих речи (кандидата за термилошки речник) из доменског корпуса из области библиотекарства и информатике

Допуна термилошког речника која се обавља поступком екстракције термина, односно одредница за речник, у нашем истраживању врши се аутоматским препознавањем вишечланих речи које су кандидати за термилошки речник из библиотекарства и информатике. Екстракција вишечланих речи изводи се на корпусу текстова из поменуте области, у овом случају то су чланци из часописа „Инфотека” и монографске публикације проф. др Александре Вранеш наведене у одељку 1.4.

Одреднице за вишечлане речи у морфолошком електронском речнику српског језика који користимо у овом истраживању, класификоване су према њиховој синтаксичкој структури, односно према информацијама које су потребне за њихову флексију. Имена класа одговарају именима коначних трансдуктора⁴²⁹ (од енг. finite state transducers – FST) који се користе за флексију вишечлане речи која припада одређеној класи. На пример, ако се вишечлана реч састоји од придева (A) иза кога следи именица (N) са којом се он слаже у погледу рода, броја, падежа и аниматности, она припада класи AXN. Символ X означава део вишечлане речи који није подложен флексији или неки сепаратор, најчешће размак или цртицу.

⁴²⁹Коначни трансдуктор је коначни аутомат са две траке – улазном и излазном. Коначни трансдуктор *трансдукује* (преводи) садржај своје улазне траке у своју излазну траку, прихватањем ниске на улазу и генерисањем нове ниске на излазу. Могућа је и ситуација да коначни трансдуктор произведе више од једног излаза за сваку улазну ниску. Коначни трансдуктори често се користе код процеса морфолошке анализе у обради природних језика.

Вишечлане речи у српском језику припадају једној од 29 различитих општих класа, али 10 од ових класа покрива више од 98% свих номиналних вишечланих речи. За екстракцију вишечланих речи које се налазе у некој од тих 10 класа развијен је посебан коначни трансдуктор који је коришћен у овом истраживању. Набројаћемо те класе које су детаљно представљене и описане у раду Крстев и сарадника⁴³⁰ (попис класа је сложен опадајуће према фреквентности – највише вишечланих речи се налази у првој, а најмање у последњој класи):

1. **AXN** – вишечлана реч се састоји од придева и именице и придев и именица се слажу у погледу све четири граматичке категорије (род, број, падеж, аниматност); пример *библиографска јединица, алфabetски каталог*;
2. **2XN** – испред именице се налази реч која не подлеже флексији; то је најчешће префикс или прилог изведен од придева, а сепаратор је најчешће цртица; пример *аудио-датотека, аудио-запис*;
3. **N2X** – иза именице се налази реч која не подлеже флексији; то је најчешће именица у генитиву или инструменталу; пример *година издавања, подручје напомена, управљање знањем*;
4. **N4X**– иза именице се налазе две речи које не подлежу флексији унутар вишечлане речи; могуће су две синтаксичке структуре:
 - а. **NNgiNgi** – иза именице се налазе два придева/именице у генитиву или инструменталу; пример *опис вишетомих публикација, препознавање именованих ентитета, штампана тачкастом матрицом, управљање изворима информација*;
 - б. **NprepNp** – иза именице се налази предлошка фраза; пример *каталогизација у публикацији, учење на даљину*;
5. **AXN2X**–испред именице се налази придев са којим се именица слаже у роду, броју, падежу и аниматности, а иза именице се налази реч која не подлеже флексији у оквиру вишечлане речи; најчешће је то именица у

⁴³⁰Cvetana Krstev, Ranka Stanković, Ivan Obradović and Biljana Lazić. „Terminology Acquisition and Description Using Lexical Resources and Local Grammars.” In: Thierry Poibeau and Pamela Faber (eds.) *Proceedings of the 11th Conference on Terminology and Artificial Intelligence, LexiCon*, Universidad de Granada, Granada, Spain, 2015, pp. 81-89, CEUR Workshop Proceedings, http://ceur-ws.org/Vol-1495/paper_13.pdf (приступљено 20.1.2016)

- генитиву или инструменталу; пример *библиографска база података, оптичко препознавање карактера*;
6. **NXN** – вишечлану реч чине две именице које се слажу у броју и падежу, а као сепаратор се најчешће јавља размак; пример *библиотека целина, рачунар домаћин*;
 7. **AXAXN** – испред именице се налазе два придева са којима се именица слаже у роду, броју, падежу и аниматности; пример *електронски информациони извори, морфолошки електронски речник*;
 8. **N6X** – изаименице се налазе три речи које не подлежу флексији унутар вишечлане речи; могуће су три синтаксичке структуре:
 - а. **NNgiPrepNp** – иза именице се налазе именица у генитиву или инструменталу и предлошка фраза; пример *конверзија текста у говор, индексирање терминима из документа*;
 - б. **NNgiNgiNgi** – иза именице се налазе три именице/придева у генитиву или инструменталу; пример *елемент података непроменљиве дужине, цифра римског бројног система*;
 - в. **NprepNpNgi** – иза именице се налази предлошка фраза; пример *податак о периодичности излажења, систем за проналажење информација*;
 9. **AXN4X** – испред именице се налази придев са којим се именица слаже у роду, броју, падежу и аниматности, а иза именице се налазе две речи које не подлежу флексији унутар вишечлане речи; могуће су две синтаксичке структуре:
 - а. **ANPrepNp** – испред именице се налази придев, а иза именице предлошка фраза; пример *Међународна организација за стандардизацију, бела књига о језицима*;
 - б. **ANNgiNgi** – испред именице се налази придев, а иза именице два придева/именице у генитиву или инструменталу; пример *Међународна федерација библиотечких асоцијација, Предметне одреднице Конгресне библиотеке*;
 10. **2XAXN** – вишечлану реч чине придев и именица који се слажу у све четири граматичке категорије, а испред њих се налази реч која не подлеже флексији унутар вишечлане речи; пример *аудио-визуелни материјали, библиотечко-информациона делатност*.

С обзиром на висок ниво хомографије у српском језику, могуће је да два или више трансдуктора препознају исту секвенцу речи од којих је само једна граматички исправна. У случају овакве двосмислености предност се увек даје случају који има већу вероватноћу појављивања према унапред дефинисаном првенству одређеног трансдуктора или фреквенције леме која је кандидат за речник. На пример, у случају двокомпонентних вишечланих речи, редослед првенства графова је следећи AXN, 2XN, N2X, NXN. Такође, код препознавања вишечланих речи тражи се најдуже могуће поклапање (сравњивање). Ако се сравњивање изврши на обрасцу AXAXN, онда се у завршној фази искључује део те вишечлане речи који је сравни обрасцем AXN.

Трансдуктори за екстракцију врше лематизацију простих речи и резултат тога не мора увек да буде исправна лема вишечлане речи. Како евалуација пронађених вишечланих речи мора бити заснована на исправним лемама, друга група коначних трансдуктора производи кандидате са исправним лемама. Те исправне леме са информацијама о њиховој синтаксичкој структури користе се за потпуно аутоматизован поступак изградње одредница за речнике у складу са стратегијом за производњу леме са свим њеним флективним облицима.^{431 432}

Неке екстраховане вишечлане речи могу имати неколико лема кандидата које су им придружене (због тога што су их препознали различити трансдуктор или због хомографије простих речи). У том случају ће бити одабрана она лема која је произведена од највећег броја различитих облика речи и највећег броја облика речи уопште узимајући у обзир фреквенцију.

Ови кораци се могу илустровати једним примером приказаним у табели 5а, а добијеним из корпуса који смо евалуирали – *библиографска јединица*. Трансдуктор који је пронашао синтаксичку конструкцију AXN (grf01) екстраховао је 6 различитих

⁴³¹ Cvetana Krstev, Ranka Stanković, Ivan Obradović, Duško Vitas and Miloš Utvić. „Automatic Construction of a Morphological Dictionary of Multi-Word Units.” In: Hrafn Loftsson, Eiríkur Rögnvaldsson, Sigrún Helgadóttir (eds.) *Proceedings of the 7th International Conference on NLP, IceTAL 2010*, Reykjavik, Iceland, August 16-18, 2010, Lecture Notes in Computer Science 6233 (Heidelberg: Springer 2010):226-237

⁴³² Ranka Stanković, Ivan Obradović, Cvetana Krstev and Duško Vitas. „Production of morphological dictionaries of multi-word units using a multipurpose tool”, In: K. Jassem, P. W. Fuglewicz, M. Piasecki and A. Przepiórkowski (eds.) *Proceedings of the Computational Linguistics-Applications Conference*, October 17–19, 2011. Jachranka, Poland, pp. 77 – 84, (Warsaw: Polish Information Processing Society, 2011)

форми – 4 које су сигурно једнина, један који је сигурно множина и један који може бити и једнина и множина.⁴³³

Табела 5а – Кораци у производњи леме за вишечлану реч

Граф	Број	Препознати облик	Фреквенција	Привремена лема	Лема
grf01	sin	bibliografska jedinica	25	bibliografski jedinica	bibliografska jedinica
		bibliografske jedinice	47		
		bibliografskoj jedinici	3		
		bibliografskom jedinicom	2		
		bibliografsku jedinicu	4		
	plu	bibliografske jedinice	47	bibliografski jedinica	bibliografske jedinice
		bibliografskih jedinica	28		
		bibliografskim jedinicama	2		

Табела 5б – Сви облици флексије екстраховане леме

DELAC	bibliografska(bibliografski.A2:aeFs1g) jedinica(jedinica.N650:fs1q),NC_AXN
DELACF	bibliografska jedinica,bibliografska jedinica.N:fs1q
	bibliografske jedinice,bibliografska jedinica.N:fs2q
	bibliografskoj jedinici,bibliografska jedinica.N:fs3q
	bibliografsku jedinicu,bibliografska jedinica.N:fs4q
	bibliografska jedinice,bibliografska jedinica.N:fs5q
	bibliografskom jedinicom,bibliografska jedinica.N:fs6q
	bibliografskoj jedinici,bibliografska jedinica.N:fs7q
	bibliografske jedinice,bibliografska jedinica.N:fp1q
	bibliografskih jedinica,bibliografska jedinica.N:fp2q
	bibliografskim jedinicama,bibliografska jedinica.N:fp3q
	bibliografske jedinice,bibliografska jedinica.N:fp4q
	bibliografske jedinice,bibliografska jedinica.N:fp5q
	bibliografskim jedinicama,bibliografska jedinica.N:fp6q
	bibliografskim jedinicama,bibliografska jedinica.N:fp7q

Привремена лема добијена лематизацијом простих речи није коректна у оба броја (једнини и множини), тако да је корекција неопходна. У фази филтрирања дошло се до тога да се задржи лема у једнини јер број различитих облика леме у једнини и фреквенција њиховог појављивања преовладавају у односу на множину.

⁴³³ Неке леме могу бити у множини тј. уопште немају једнину иако компоненте леме могу имати облик у једнини. На пример, такве леме су *допунски елементи*, *каталошка правила*, *људски ресурси*, *подаци о аутору*, *техничке могућности* и др.

Овај облик леме са информацијама о њеној структури довољан је за аутоматску производњу леме за електронски речник (DELAC), док информације о лемима омогућавају даљу производњу свих облика флексије (DELACF) (види табелу 5б).

Рангирање екстрахованих вишечланих речи (кандидата за термине)

Скуп лема који је добијен поменутом процедуром даље се филтрира увођењем мера које комбинују језичке и статистичке информације. За сваку лему, поред фреквенције, прорачунавају се основне мере (C-Value, NC-Value, T-Score, LLR и Keynes) и оне мере које комбинују основне мере. На основу одабране мере и адекватних вредности, генерише се скуп лема за евалуацију. Резултати мера фреквенције, C-Value, NC-Value⁴³⁴ и T-Score показују ранг (позицију) леме у оквиру доменског корпуса, а две преостале мере (LLR^{435, 436} и Keynes⁴³⁷) пореде фреквенцију термина у доменском корпусу и у општем језичком корпусу и на тај начин указују колико је одређена вишечлана реч специфична за изабрани домен. Аутоматски прорачунате мере и ранг изабране вишечлане речи, односно термина (нпр. библиографска јединица) заједно са другим терминима приказани су у табели 6.

Објаснићемо значења мера које су коришћене у истраживању. Вредност C-Value повећава општу статистичку меру фреквенције појављивања у екстракцији термина утичући на проналажење специфичних вишечланих речи, нарочито у случају речи које су садржане у неким другим речима (лексичка гнезда). Приметићемо да је у табели 6 вредност C-Value приближна или иста као и мера фреквенције осим у примерима *мрежа речи*, *семантичка мрежа*, *семантичка мрежа речи* јер је рангирање ниже код термина који су садржани у неком другом термину. NC-Value је вредност која даје метод за екстракцију термина у контексту (речи које имају тенденцију да се појављују у оквиру термина) и укључује информације из терминолошки оријентисаних речи у екстракцију нових термина. T-Score је мера која

⁴³⁴ Katerina Frantzi, Sophia Ananiadou and Hideki Mima. „Automatic recognition of multi-word terms: the C-value/NC-value method.” *International Journal on Digital Libraries* 3,2(2000):115-130.

⁴³⁵ Ted Dunning. „Accurate methods for the statistics of surprise and coincidence.” *Computation Linguistics* 19,1(1993):61-74.

⁴³⁶ Alexander Gelbukh, Grigori Sidorov, Eduardo Lavin-Villa and Liliana Chanona-Hernandez. „Automatic Term Extraction Using Log-Likelihood Based Comparison with General Reference Corpus.” In C. Hopfe, Y. Rezgui, E. Métais, A. Preece & H. Li (eds.), *Natural Language Processing and Information Systems*, Vol. 6177, pp. 248-255. (Berlin, Heidelberg: Springer, 2010).

⁴³⁷ „Statistics used in Sketch Engine.” *Lexical Computing*, 2015, <https://www.sketchengine.co.uk/wp-content/uploads/ske-stat.pdf> (приступљено 17.2.2016)

се односи на мерење адхезије (веза) између две речи у корпусу. LLR (Log-Likelihood Ratio) мера се у овом случају односи на разлике између доменског корпуса и општег корпуса савременог српског језика, али степен реакције на те разлике варира.

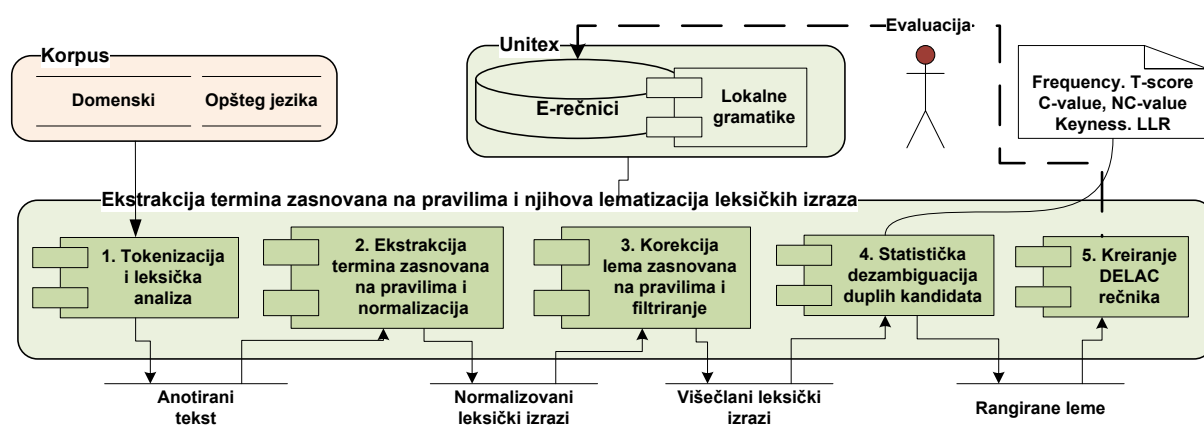
Keyness је статистички индекс који се користи за евалуацију у смислу процене колико је нека реч значајна за одређени корпус или документ. У примеру *библиографска јединица* у нашој табели 6 вредност Keyness опада јер је термин релативно мање фреквентан у доменском корпусу него у општем корпусу савременог српског језика. Зато, ако погледамо табелу 6, уочићемо да је за лему *српски језик* вредност Keyness изузетно висок јер се тај термин чешће среће у општем корпусу него у доменском корпусу. Обрнута ситуација је код леме *универзитетска библиотека*. Вредност Keyness је ниска јер се ова лема много чешће јавља у доменском корпусу него у општем корпусу српског језика.

Табела 6 – Рангирање пронађених облика и лема вишечланих речи

graf	broj	lema	eliminisani kandidati	Vrednosti mera						Rangiranje					
				Frek	Cval	TScr	LLR	Kness	LCVal	Frek	Cval	TScr	LLR	Kness	LCVal
grf01	sin	univerzitetska biblioteka	univerzitetske biblioteke	373	361	19	2549	226	1455	1	1	1	1	1	1
grf01	sin	sрpski јezik		205	205	14	366	6	285	2	2	2	17	834	11
grf01	sin	narodna biblioteka	narodne biblioteke	192	192	14	749	21	471	3	3	3	4	220	3
grf03	sin	baza podataka	baze podataka	177	168	13	924	49	546	4	4	4	2	56	2
grf01	sin	bibliotečka града	bibliotečke граде	95	95	10	681	151	388	8	7	7	6	2	6
grf01	sin	filološki fakultet		88	75	9	309	16	192	9	12	9	23	436	21
grf01	sin	bibliografski opis		87	79	9	612	131	345	10	10	10	8	6	8
grf01	sin	biблиографска јединица	библиографске јединице	81	81	9	570	124	325	11	9	11	9	7	9
grf01	plu	jezički resursi		77	77	9	541	120	309	12	11	12	10	8	10
grf01	sin	informacioni sistem		76	56	9	287	18	172	13	20	13	32	336	28
grf01	sin	bibliotečki materijal		70	70	8	492	111	281	14	13	14	11	9	12
grf03	sin	biblioteka univerziteta	biblioteke univerziteta	68	50	8	478	109	264	15	25	16	12	10	13
grf01	sin	svetski rat		68	68	8	6	1	37	16	14	15	906	934	315
grf01	sin	kongresna biblioteka	kongresne biblioteke	41	34	6	288	73	161	44	65	45	31	24	33
grf01	sin	bibliotečko poslovanje		41	41	6	288	73	165	45	45	43	30	23	30
grf01	sin	arhivska града	arhivske граде	40	40	6	151	17	96	46	46	46	87	434	73
grf01	sin	људско зnanje		28	28	5	159	34	93	81	85	80	77	99	76
grf01	sin	научна област	научне области	28	28	5	143	28	86	82	86	83	97	150	93
grf01	sin	radna grupa	radne grupe	28	22	5	22	3	22	83	125	82	825	898	744
grf04a	sin	program za etiketiranje		28	44	5	197	53	121	84	36	81	59	50	54
grf01	sin	računarska lingvistika	računarske lingvistike	27	27	5	190	51	108	85	93	86	62	51	61
grf01	sin	nabavna politika	nabavne politike	27	27	5	190	51	108	86	91	85	63	52	62
grf01	plu	informacioni sistemi		27	19	5	98	14	58	87	149	88	176	546	162
grf03	sin	fond biblioteke		27	27	5	190	51	108	88	88	92	65	54	64
grf01	plu	bibliotečki fondovi		27	27	5	190	51	108	89	90	87	64	53	63
grf03	sin	mreža reči		26	18	5	183	49	100	90	160	89	68	55	69
grf01	sin	semantička mreža	semantičke mreže	21	13	5	148	41	80	120	257	126	94	79	99
grf05	sin	semantička mreža reči	semantičke mreže reči	8	13	3	56	17	34	585	269	570	394	360	381

Архитектура система

Аутоматска процедура за екстракцију вишечланих речи као кандидата за термине и лематизација остварена је коришћењем система Unitex за обраду корпуса и LeXimir-a, вишенаменског софтверског алата за управљање лексичким ресурсима. Целокупан процес је аутоматизован почевши од токенизације и лексичке анализе текста до производње одредница за речнике, а потребна је само незнатна ручна интервенција. Систем се ослања на позитивне карактеристике Unitex-а за анализу текста и израду коначних трансдуктора, као и на бројне функционалности LeXimir-а у производњи лема у стандардним форматима. Сви резултати и припадајући метаподаци похрањени су у бази податакана SQL серверу.⁴³⁸



Слика 71 – Архитектура система

Архитектура софтверског решења приказана је на слици 71, а заснована је на веб сервисима, па омогућава другим апликацијама да користе неке од ових сервиса (индексирање или проналажење информација у документу) за потребе екстракције термина. Апликација која се тренутно користи развијана је и тестирана у окружењу Windows док је нова веб апликација тренутно у изради, а она треба да понуди екстракцију термина из различитих домена ширем кругу корисника.⁴³⁹

Евалуација

За евалуацију смо користили корпус који се састоји од 4 књиге и 72 чланка из часописа „Инфотека” из области библиотекарства и информатике. Величина тог

⁴³⁸ SQL (Structured Query Language) server, доступно на: <http://www.microsoft.com/en-us/server-cloud/products/sql-server/>

⁴³⁹ Ranka Stanković, Cvetana Krstev, Ivan Obradović, Biljana Lazić and Aleksandra Trtovac. „Rule based automatic multi-word term extraction and lemmatization.” In: *10th edition of the Language Resources and Evaluation Conference (LREC)*, 23-28 May 2016, Portorož (Slovenia) (у штампи)

корпуса је 22.948 реченица и 480.588 облика простих речи. За прорачунавање одређених вредности било је потребно поредити резултате са општим корпусом и у те сврхе смо користили 22 милиона речи из Корпуса савременог српског језика.

Нашом претрагом добијено је 89.195 облика вишечланих речи са укупно 56.343 лема. За евалуацију смо задржали само оне леме са фреквенцијом већом од 3 појављивања у корпусу. Подела према трансдукторима помоћу којих су пронађене оне вишечлане речи и леме које смо задржали приказана је у табели 7. Од 2.655 лема, одбачено је 72 дупликата, и 2.583 лема покрива 24.699 препознатих облика који су задржани. За те облике прорачунате су мере и они су према њима рангирани и прослеђени на евалуацију.

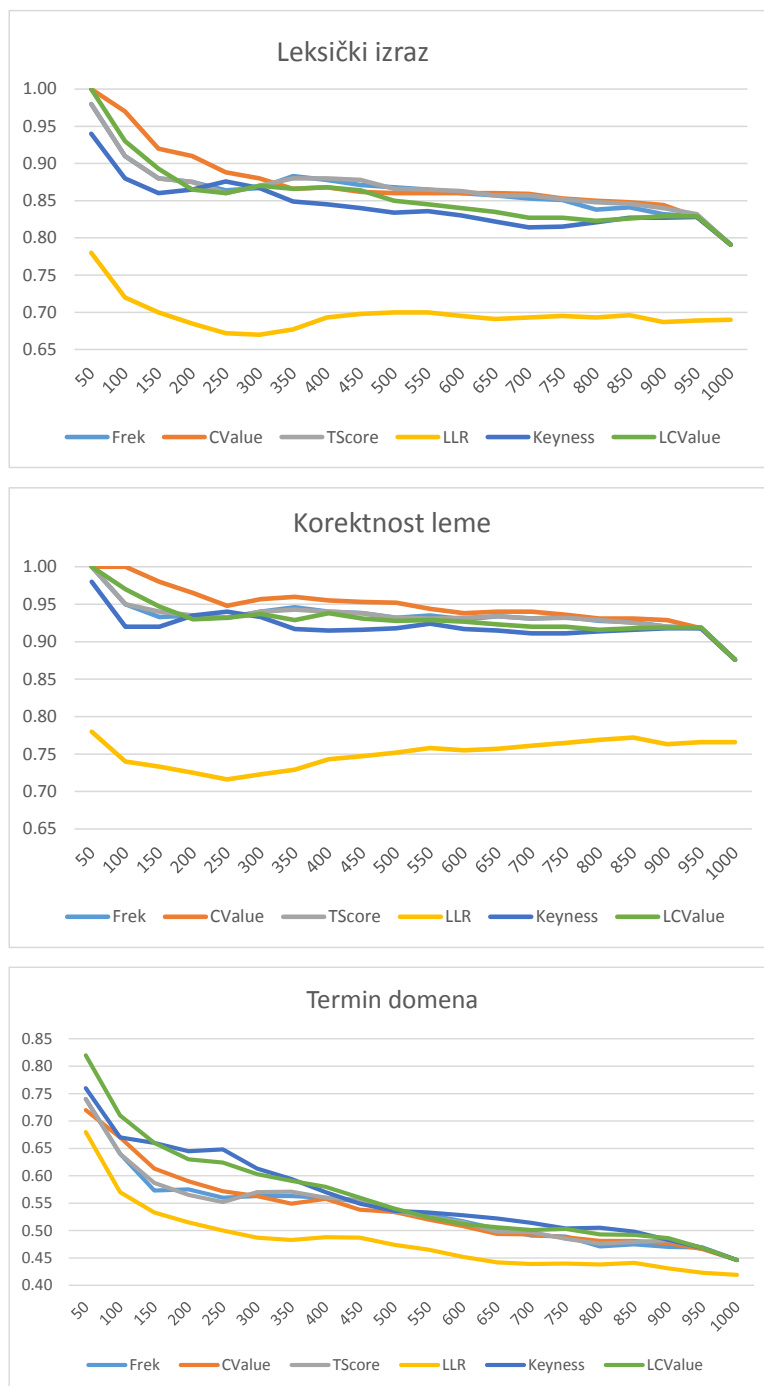
Табела 7 – Број лема са више од 3 појављивања у доменском корпусу

Graf				Množina (plu)		Jednina (sin)	
				Lema	Oblika	Lema	Oblika
grf01	AXN	AN	AN	438	5398	611	7825
grf02	2XN	(An ADV Pref !SDIC)[-]?N	X-N	5	58	12	161
grf03	N2X	Nngi	NNgi	212	1810	462	3655
grf04a	N4X	NprepNp	NprepNp	123	843	162	1093
grf04б	N4X	N((A N)gNg (A N)iNi)	NNgiNgi	36	196	89	561
grf05	AXN2X	ANNgi	ANNgi	17	106	54	360
grf06	NXN	N[-]?N	N-N	163	1240	193	1478
grf07	AXAXN	AAN	AAN	4	26	14	102
grf08a	N6X	NngiPrepNp	NNgiPrepNp	3	14	3	16
grf08б	N6X	N((A N)g(A N)gNg (A N)i(A N)iNi)	NNgiNgiNgi			1	4
grf08в	N6X	NprepNpNgi	NPrepNpNgi	12	85	14	99
grf10	2XAXN	(An ADV Pref !SDIC)[-]?AN	X-AN	9	63	18	167
Ukupno				1022	9839	1633	15521

Ukupno (plu+sin)				2655	25360
------------------	--	--	--	------	-------

Процес евалуације тече у три фазе: 1) за сваки кориговани облик потребно је одлучити да ли је вишечлана реч исправна (нпр. исправна вишечлана реч *народна поезија*, неисправна вишечлана реч *књига библиотеке*), 2) за сваки предложени облик леме треба утврдити да ли је лема тачна (нпр. исправна лема *ћурилично писмо*, неисправна лема *база знање*) и 3) за сваку исправну вишечлану реч треба одлучити да ли је одговарајућа као термин за терминолошки речник из библиотекарства и информатике (нпр. термин је *библиографски опис*, термин није *евалуација у библиотекама*). Прецизност проналажења израчуната је за сваки од ових корака, а

резултати су представљени на слици 72 за групе од по неколико стотина ранжираних вишечланих речи према основним мерама: фреквенција, C-Value, T-Score и Keynes и комбинацијом основних мера $TKValue = T-Score * Keynes$. Средња просечна прецизност дата на дну сваке од слика показује да су све мере дале сличне резултате.



Слика 72 – Прецизност проналажења 1) вишечлане речи; 2) леме; 3) термина

У даљем раду на допуни овог терминолошког речника, намера је да се користи веб апликација која је тренутно у изради, као и да се повећа прецизност препознавања исправне леме развојем нових коначних трансдуктора за екстракцију потенцијалних

термина. Такође, потребно је проверити и евалуирати садашње листе лема са фреквенцијом појављивања у корпусу мањом од три јер се и у том делу екстрахованих лема могу наћи кандидати за термилошки речник (пример *библиографска референца уз текст, библиографски опис на вебу, биографски податак о аутору, графички запис са камена, Међународна федерација за документацију, напредна машина за претраживање* и сл.)

6.2.3 Анализа резултата у допуни термилошког и општег речника српског језика

Допуна термилошког речника из библиотекарства и информатике

У претходном поглављу описали смо процес евалуације у три фазе. У овом поглављу детаљно ћемо описати практични рад на евалуацији поткрепљен примерима. Кроз ова три корака евалуације, објаснићемо на који начин смо донели одлуку о изабору термина за допуну термилошког речника из библиотекарства и информатике.

У првој фази евалуирали смо препознате вишечлане леме у смислу исправности, односно утврђивања да ли је препозната ниска заиста вишечлана реч (MWU). Током евалуације одбацили смо 378 кандидата. Код 150 кандидата од поменутих 378 утврђено је да лема није тачна, а та анализа ће бити описана у даљем тексту. У неким случајевима одбацивање је извршено због погрешног броја именице (уместо множине јавља се једнина – *постдипломска студија, докторска студија*; уместо једнине, јавља се множина – *име писаца, критички осврти*). Од укупног броја одбачених вишечланих речи, 244 кандидата је одбачено зато што не носе специфично значење или је у питању непотпун назив, односно назив који је садржан у другој вишечланој речи. Навешћемо примере таквих случајева са највећом фреквенцијом појављивања у доменском корпусу⁴⁴⁰: *факултет универзитета* (38), *библиотека републике* (34), *српски семинар* (33), *српска академија* (26), *скупштина заједнице* (23), *библиотека матице* (21), *летопис матице* (18), *републички завод* (17), *тип библиотека* (16), *група за језичке* (15), *међународна федерација* (13), *систем универзитета* (12), *јануар-фебруар* (10) итд.

У другој фази евалуације испитали смо тачност предложених лема. Као што смо већ навели, у 150 случајева лема није била тачна. Највећи број неисправних лема (106)

⁴⁴⁰Фреквенција је наведена у загради

је код потенцијалних вишечланих речи које су препознате помоћу обрасца NXN (трансдуктор 06) који дефинише да вишечлану реч чине две именице које се слажу у броју и падежу, а као сепаратор се најчешће јавља размак. Следећи примери илуструју поменути случај: *библиотека универзитет* (31), *академија наук* (29), *облик реч* (25), *факултет универзитет* (24), *врста реч* (22), *милион реч* (19), *систем квалитет* (17), *скупштина заједница* (16), *библиографија библиографија* (14), *наслов часопис* (12), *претраживање информација* (10) итд. Иако би примери *библиографија библиографија* и *претраживање информација* били кандидати за термилошки речник, они у овом случају нису одабрани пошто је обрасцем NXN дефинисано да су обе именице у истом падежу, у овом случају номинативу који је у наведеним примерима за другу именицу у вишечлавној речи једнак генитиву множине. Исправна лема би требало да садржи управо генитив множине. Такође, занимљив пример је и *академија наук*, где је уместо именице *наука*, за лему понуђена именица *наук*.

Екстракција вишечланих речи преко обрасца N2X (иза именице се налази реч која не подлеже флексији; то је најчешће именица у генитиву или инструменталу) дала је 26 нетачних лема: *фонду библиотеке* (19), *приказа садржаја* (21), *ранг листи* (10), *веб апликације* (14), *веб сајта* (10) итд. Уместо именице *фонд*, погрешно је као део леме предложена именица *фонду*, а уместо именице *приказ*, именица *приказа*. Иако би се у случају леме *веб апликације* могло сматрати да се ради о множини, образац јасно дефинише да је други део вишечлане речи реч која не подлеже флексији, у овом случају именица у генитиву, тј. Ову вишечлану реч исправно препознаје образац 2XN.

Образац N4X је из доменског корпуса екстраховао вишечлане речи за које се јавља 10 неисправних лема: *библиотека у наставу* (4), *фонду универзитетске библиотеке* (5), *листа за књижевност* (4), *листа за забаву* (6), *наук о земљи* (5) и сл. Анализом наведених примера можемо утврдити да је именица *фонд* поново погрешно препозната као именица *фонду*, док је именица *лист* замењена именицом *листа*. Такође, опет је уместо именице *наука* понуђена именица *наук*.

Претрага путем обрасца AXN (вишечлана реч се састоји од придева и именице и придев и именица се слажу у погледу све четири граматичке категорије) резултирала је са 6 неисправних лема, од којих су најзанимљивији примери: *библиотечки фонду* (25), *књижни фонду* (6), *завичајни фонду* (10). У сва три случаја, уместо именице *фонд*, у лемама се погрешно јавља именица *фонду*.⁴⁴¹

⁴⁴¹Фонду – назив за швајцарско национално јело од топљеног сира

Образац 2XN дао је једну неисправну лему – *библио-тек* (4), као и образац AXN2X – *српско женска имена* (4). У првом случају, реч је о хифенацији на крају реда која је погрешно препозната као сепаратор,⁴⁴² а у другом случају о погрешно сложеном броју и роду придева и именице.

Потребно је посебно напоменути да ниједна од препознатих вишечланих речи није остала без исправне леме. Исправну лему је понудио трансдуктор неког другог обрасца.

Последња, трећа, фаза евалуације односи се на одабир термина за термилошки речник. Уколико је вишечлана реч била ниска речи која носи специфично значење и ако је лема била коректно препозната, онда смо приступили одабиру термина за термилошки речник. Ако размотримо евалуацију термина за речник код термина који имају фреквенцију већу од 3 појављивања у доменском корпусу, уочићемо да је од укупног броја од 1.426 коректно препознатих вишечланих речи одабрано 592 кандидата (термин) за термилошки речник. Целокупан попис термина одабраних за термилошки речник из библиотекарства и информатике дат је у прилогу 1 у абecedном поретку.

Ако погледамо прилог 1 видећемо да се одређене леме издвајају по својој фреквенцији појављивања у доменском корпусу. У табели 8 дали смо листу 25 најфреквентнијих лема.

Прве три леме јављају се са фреквенцијом већом од 150 појављивања, па ћемо детаљније објаснити који су разлози за ову појаву.

Уочљиво је да је лема са убедљиво највећом фреквенцијом од 373 појављивања у доменском корпусу *универзитетска библиотека*. Разлог за то лежи у чињеници да се велики број текстова из часописа „Инфотека” у доменском корпусу односи на функционисање и иновативне процесе у универзитетским библиотекама, а целокупан специјални број из 2004. године на развој нових библиотечких сервиса и услуга у Универзитетској библиотеци „Светозар Марковић”у Београду, Универзитетској библиотеци „Никола Тесла” у Нишу и Универзитетској библиотеци у Крагујевцу и поређењу са стањем у другим универзитетским библиотекама у Европи. Такође, тема једне од монографских публикација др Александре Вранеш која је део доменског корпуса су високошколске библиотеке у које свакако спадају и универзитетске

⁴⁴² Цртица је остала у тексту услед грешке приликом аутоматског претварања текстова из PDF формата у чист текст

библиотеке, па и ту лежи објашњење за овако високу фреквенцију поменуте леме. Оба ова случаја објашњавају и високу фреквенцију леме *високошколска библиотека* (95).

Табела 8 – Најфреквентније вишечлане леме у доменском корпусу

Образас	graf	broj	Lema	FrekGrafLema
AXN	grf01	Sin	univerzitetska библиотека	373
AXN	grf01	Sin	narodna библиотека	192
N2X	grf03	Sin	baza podataka	177
AXN	grf01	Sin	bibliotečka delatnost	107
AXN	grf01	Sin	srpska bibliografija	100
AXN	grf01	Sin	visokoškolska библиотека	95
AXN	grf01	Sin	bibliografski opis	87
AXN	grf01	Sin	bibliografska jedinica	81
AXN	grf01	Sin	informacioni sistem	76
AXN	grf01	Sin	bibliotečki materijal	70
AXN	grf01	Sin	autorsko pravo	65
AXN	grf01	Sin	naučni rad	62
AXN	grf01	Sin	međubibliotečka pozajmica	61
N2X	grf03	Sin	ime autora	56
N2X	grf03	Sin	pronalaženje informacija	52
AXN	grf01	Sin	grupna raspodela	51
AXN	grf01	Sin	izdavačka produkcija	50
AXN	grf01	Sin	ključna reč	48
2XAXN	grf10	Sin	bibliotečko-informacioni sistem	47
AXN	grf01	Plu	bibliografski podaci	46
AXN	grf01	Sin	zavičajna bibliografija	44
AXN	grf01	Sin	elektronsko učenje	43
AXN	grf01	Sin	istraživački rad	42
AXN	grf01	Sin	kongresna библиотека	41
AXN	grf01	Sin	arhivska građa	40

Ако погледамо целокупну листу препознатих лема, приметимо да се не појављује лема *Универзитетска библиотека „Светозар Марковић”* иако се анализом доменског корпуса може установити њено вишеструко појављивање. Само у корпусу часописа „Инфотека”, претрагом коју смо извршили у оквиру алата *Библиша*, добили смо резултат да се ова лема јавља у 22 конкорданце.

Разлог због ког ова лема није издвојена из доменског корпуса је то што ниједан трансдуктор не само да није предвидео структуру према следећем обрасцу – именица испред које је придев са којим се она слаже у падежу, роду, броју и аниматности, а иза које су једна или више именица у номинативу при чему су то властите именице, већ је експлицитно дефинисано да властита именица не може стајати иза придева и именице.

Лема која има високу фреквенцију у доменском корпусу је и *народна библиотека* са фреквенцијом 192. Међутим, анализом и претрагом доменског корпуса и у испису конкорданци у оквиру алата *Библиша* запажамо да је у великом броју случајева реч о лема *Народна библиотека Србије* чији је саставни део лема *народна библиотека*. У следећем кораку истражили смо зашто у анализираном корпусу применом графова и образаца није препозната и из њега издвојена лема *Народна библиотека Србије*. Ова конструкција вишечланих речи препознаје се и издваја помоћу трансдуктора 5, односно помоћу обрасца AXN2X којим је дефинисано да се испред именице налази придев са којим се именица слаже у роду, броју, падежу и аниматности, а иза именице се налази реч која не подлеже флексији у оквиру вишечлане речи, најчешће је то именица у генитиву или инструменталу. Према овако дефинисаном обрасцу, лема *Народна библиотека Србије* морала би неизоставно да буде препозната. Ипак, трансдуктор 5 је дефинисан тако да трећа реч не може бити властито име. Управо је то разлог што поменута лема није препозната.

Потребно је појаснити да је примарни разлог за то што леме ове врсте нису издвојене из доменског корпуса то што називи институција какви су *Универзитетска библиотека „Светозар Марковић“* и *Народна библиотека Србије* нису термини у ужем смислу, иако јесу вишечлане речи. Препознавање назива институција у корпусу текстова и њихово издвајање врши се алатима за препознавање именованих ентитета (NER), па је то био основни разлог да графови искључе проналажење лема ове врсте.

Лема база података бележи 177 појављивања у доменском корпусу. Пошто данас скоро свака библиотека има бар базу података о свом фонду, онда не чуди многострука употреба овог термина у научним текстовима из анализираних области. Овај термин је осим широке употребе у подручју библиотекарства и информатике, ушао и у свакодневну употребу код говорника српског језика.

Из пописа најфреквентнијих лема у доменском корпусу запажамо да су неке од њих, као што је *база података*, у широкој употреби и у свакодневној комуникацији, па је битно нагласити да су 53 леме екстраховане из доменског корпуса већ постојале у речнику DELAC. Издвојићемо неке од лема које се већ налазе у речнику DELAC, а имају фреквенцију појављивања већу од 20: *база података* (177), *информациони систем* (76), *ауторско право* (65), *насловна страна* (43), *програмски језик* (42), *радно време* (34), *информациона технологија* (28), *информациона делатност* (24), *децимална класификација* (22), *обрада података* (22), *софтверска пиратерија* (22),

веб локација (22) итд. У прилогу 2а дата је комплетна листа тих лема са фреквенцијом њиховог појављивања.

Важно је објаснити и зашто су одређене вишечлане речи одбачене и искључене као кандидати за терминолошки речник. Неколико разлога је довело до такве ситуације:

- Термини су општег значења – такви термини ће бити уврштени у општи речник српског језика (објашњења са примерима биће наведена у наредном одељку);
- Лема за вишечлану реч се јавља у множини, а термин у речнику треба да буде у једнини;

Комплетна листа лема ове врсте садржи 90 вишечланих лема и дата је у прилогу 2б. На овом месту ћемо, као илустрацију ове појаве, навести неке од најфреквентнијих лема: *библиографски записи* (34), *библиотечки фондови* (27), *информациони центри* (11), *информациони извори* (29), *информациони системи* (31), *истраживачке библиотеке* (10), *међународни стандарди* (35), *новински чланци* (10), *радни процеси* (15), *саставни делови* (11), *завичајни фондови* (20) итд.

Поједине вишечлане леме које су екстраховане из доменског корпуса у множини, такође су екстраховане и у облику у једнини, па је тај облик ушао у терминолошки речник. Такви су термини: *библиографски опис* (87), *библиотечки фонд* (33), *библиотечки систем* (42), *библиотечко-информациони систем* (47), *централни каталог* (39), *саставни део* (34), *завичајни фонд* (15) и др. Целокупан попис од 22 овакве леме дат је у прилогу 2в.

Ако погледамо ове две листе речи видећемо да је фреквенција појављивања у неким случајевима знатно већа код облика у множини који је одбачен него код облика у једнини (нпр. *библиографски записи* (34) – *библиографски запис* (8); *каталошки листићи* (29) – *каталошки листић* (6); *међународни стандарди* (35) – *међународни стандард* (20)). Закључак је да висока фреквенција није једини предуслов да би једна вишечлана лема постала кандидат за термин у речнику.

Запазићемо да је разлика између лема које се јављају само у множини и оних које се јављају и у једнини 64. Након аутоматске екстракције ручном евалуацијом је предложено да уместо облика у множини облик у једнини буде облик за речник.

- Вишечлана лема се јавља у једнини, а термин у речнику треба да буде у множини;

Погледаћемо листу од 9 вишечланих лема које се јављају у једнини, а за речник је потребан термин у множини: *библиографски преглед* (10), *библиотечка услуга* (14), *информациона наука* (10), *информационо-комуникациона технологија* (4), *језичка технологија* (43), *комуникациона технологија* (8), *научна информација* (12), *потреба корисника* (35), *рачунарска наука* (8).

Једна од ових речи је екстрахована из доменског корпуса и у облику у множини, па је тај облик одабран за терминолошки речник – *библиографски прегледи* (9).

- Вишечлана лема је исправна и реч терминолошки припада речнику, али је други термин, синоним, одабран за речник;

Ова појава уочљива је у следећим случајевима: *библиографски опис публикација* (4) – одабран је термин *библиографски опис* (87), *централни каталог грађе* (4) – одабран је термин *централни каталог* (39), *посебна библиотека целина* (5) – одабран је термин *библиотека целина* (12), *рачунарска лингвистика* (6) – одабран је термин *рачунарска лингвистика* (27), *распоређивање грађе* (5) – одабран је термин *распоред грађе* (4), *саставни делови публикација* (5) – одабран је термин *саставни део* (34).

Међутим, јавља се и друга ситуација – и вишечлана реч и њен синоним одабрани су као термини за речник. Илустроваћемо ову појаву неколицином примера: *библиографско описивање* (22) и *библиографски опис* (87), *библиотека универзитета* (68) и *универзитетска библиотека* (373), *фонд библиотеке* (27) и *библиотечки фонд* (33), *књига инвентара* (4) и *инвентарска књига* (18) и др. За целокупну листу оваквих термина погледати прилог 2г.

- Вишечлана лема представља назив часописа или назив институције – то није термин у ужем смислу;

У току претраге из доменског корпуса аутоматски су екстраховане и следеће вишечлане речи које представљају називе листова и часописа или потпуне или скраћене називе институција. Навешћемо неке од наслова периодичних публикација: *Бачка вила* (7), *Босанска вила* (5), *Јужнословенски филолог* (18), *Крфски забавник* (6), *Српски народни лист* (5), *Статистички годишњак Републичког завода* (5) и др, као и неке од институција чији су називи екстраховани из доменског корпуса: *Филолошки факултет* (88), *Филозофски факултет* (62), *Математички факултет* (44), *Музеј позоришне уметности* (5), *Позориште на Теразијама* (5), *Српска академија наука* (25) и др.

Како називи часописа и називи институција нису термини, одлучено је да ове вишечлане речи буду искључене као кандидати за терминолошки речник.

Допуна општег речника српског језика

Ипак, пошто један доменски корпус није сачињен само од стручне терминологије, аутоматском екстракцијом вишечланих лема, што је видљиво и из претходног дела, нису добијени само кандидати за термилошки речник. Анализом лема, установили смо да се јављају две ситуације:

- 1) лема је кандидат за општи речник српског језика;
- 2) лема је кандидат и за термилошки речник из библиотекарства и информатике и за општи речник српског језика.

1)

У првом случају, установили смо да се из доменског корпуса међу екстрахованим и исправно препознатим лемама налази и 129 кандидата за општи речник српског језика.

Неки од кандидата за општи речник српског језика аутоматски екстраховани из доменског корпуса библиотекарства и информатике су: *академско образовање, балкански језици, цивилно становништво, драмски писац, државна институција, финансијска подршка, хуманистичке науке, историјска вредност, језичка баријера, критички осврт, куповна моћ, људска права, научна комуникација, позоришна уметност, правна заштита, ранг листа, студијско путовање, установа културе, високошколска институција, заштита природе* итд. Потпуна листа ових лема налази се у прилогу 3а.

Као и у случају одабира кандидата за термилошки речник, и код одабира термина за општи речник српског језика наишли смо на ситуацију да је екстрахована лема у множини, а да је за речник потребна лема у једнини. Таква појава се јавила у 35 случајева. Навешћемо неке од примера: *наставни планови (15), наставни програми (20), научни радови (15), позоришне представе (9), предшколске установе (6), телевизијске емисије (8), универзитетски професори (4), временски оквири (8), законски прописи (6)*. Пуна листа лема ове врсте дата је у прилогу 3б.

Такође, јављају се и случајеви да је екстрахована лема у једнини, а термин треба да буде у множини: *докторска студија (5), хуманистичка наука (12), постдипломска студија (5)*.

2)

У другом случају утврдили смо да постоји 94 кандидата за које сматрамо да треба да буду део и термилошког и општег речника српског језика. Одабир ових речи заснован је на критеријуму да се те речи, осим у области библиотекарства и

информатике, користе и у другим научним областима или у општој комуникацији. Леме за ове вишечлане речи нису се налазиле у DELAC-у пре овог истраживања.

Навешћемо неке од термина који су одабрани и за термилошки речник библиотекарства и информатике и за општи речник српског језика: *циљна група*, *географска карта*, *радне обавезе*, *радни простор*, *радno место*, *радno окружење*, *статистички метод*, *студија случаја*, *техничка документација*, *техничка подршка*, *техничке могућности*, *управљачка структура*, *веишачка интелигенција*, *законска регулатива*, *заштита споменика културе* и др. Целокупан попис лема ове врсте дат је у прилогу 4.

На основу анализе ове листе речи можемо утврдити да оне у оквиру термилошког речника у великој мери спадају у област библиотечког пословања, а у оквиру општег речника у област пословања и управљања (*кадровска политика* (6), *контрола квалитета* (7), *људски ресурси* (6), *набавна политика* (27) *надзорни одбор* (12), *организациона јединица* (10), *организациона структура* (11), *правни статус* (7), *програмски савет* (8), *радна група* (28), *радна организација* (6), *радno место* (12), *стратешки план* (11), *техничка подршка* (8), *тимски рад* (8), *управљачка структура* (4), *законска регулатива* (10) итд.).

Издвојићемо и примере за термине који се односе на називе писама (*ћирилично писмо* (6), *грађанска ћирилица* (4), *латинично писмо*(8), *словенско писмо* (10)) или термине везане за информатику и интернет (*дискусиона група* (5), *информациона писменост* (11), *кључна реч* (48), *корисничко име* (7), *рачунарска технологија* (17), *учење на даљину* (32), *веб сајт* (5), *веишачка интелигенција* (4) итд.).

Код термина који су сачињени од именице којој претходи придев *међународни,-а,-о* са којим се именица слаже у свим граматичким категоријама, за оба речника одабрали смо следеће термине: *међународна конференција* (13), *међународна организација* (14), *међународна сарадња* (11), *међународни карактер* (8), *међународни стандард* (29). Објаснићемо разлоге за овакав одабир. Међународне конференције организују се у свим областима знања, међународне организације регулишу правила и принципе рада у различитим областима, међународна сарадња постоји како између библиотека, тако и између других установа, међународни карактер је тежња у највећем броју достигнућа, а међународни стандард императив у науци и техници.

После завршеног избора речи како за термилошки речник тако и за општи речник српског језика, приступили смо анализи новог текста на коме смо применили

нове речнике који су допуњени одабраним терминима, као и постојеће графове за екстракцију вишечланих лема. Резултати ове анализе описани су у следећем одељку.

Примена алата за препознавање именованих ентитета (NER) у доменском корпусу

Доменски корпус је, као што смо делимично видели из анализе екстрахованих вишечланих речи, богат појмовима који су везани за препознавање именованих ентитета.

Издвојени именовани ентитети подељени су у неколико основних категорија (особа, организација, локација, мера, време), са поткатегоријама. За сваку од категорија и поткатегорија, илустрације ради, навешћемо по три примера који се у листи именованих ентитета јављају са највећом фреквенцијом⁴⁴³:

- особа – име уважене личности (семантички маркер <pers.spec>
кнеза Милоша (6), патријарха Макарија (4), архимандрит Леонид (3)
- особа – име и презиме односно пуно име (семантички маркер <persName.full>
Светозар Марковић (99), Стојана Новаковића (69), Павле Поповић (59)
- особа – улога, односно појединачна функција (семантички маркер <role>
професор (30), уредник (20), аутор (19)
- особа – лично име (семантички маркер <persName.first>
Милош (6), Макари (4), Леонид (4)
- особа – презиме (семантички маркер <persName.last>
Поповић (65), Новаковић (60), Шафарик (43)
- организација – комерцијалне организације на страном језику (семантички маркер <org.com>
IFLA Council (9), International Federation of Library Association (6), IMTEL Computers (6)
- организација – политичке организације (семантички маркер <org.pol>
Балкански савез (1), Напредне странке (1), Народне странке (1)
- организација – друге организације (семантички маркер <org>
библиотека (313), универзитет (115), заједница (108)
- локација – држава (семантички маркер <top.dr>
Србија (589), Југославија (101), САД (81)
- локација – насељено место (семантички маркер <top.gr>

⁴⁴³ За неке од поткатегорија препознато је мање од три именована ентитета

- Београд (582), Нови Сад (92), Беч (55)*
- локација – водена површина (семантички маркер <top.hyd>)
Амазон (22), Ин (13), Охајо (13)
 - локација – географска површина (семантички маркер <top.geo>)
Јавор (31), Стражилово (20), Цер (4)
 - локација – регија унутар државе (семантички маркер <top.reg>)
Република Српска (25), Војводина (22), Град Београд (19)
 - локација – део града (семантички маркер <top.deoGr>)
Будим (9), Земун (9), Пешта (8)
 - локација – регија већа од једне државе (семантички маркер <top.supReg>)
Европа (95), Америка (18), Северна Америка (11)
 - мера – тачно одређена мера (семантички маркер <measure.exact>)
30 дана (5), 45ст (4), 7,5h (3)
 - мера – приближна мера (семантички маркер <measure.approx>)
око 300g (1), око 2,5тт (1), око 5тт (1)
 - мера – опсег мера (семантички маркер <measure.range>)
9-18h (1), 9-15h (1), 5-6 сати (1)
 - мера – тачно одређен новчани износ (семантички маркер <money.exact>)
100.000 долара (2), 267.210 динара (2), 1.143.000 динара (2)
 - мера – приближан новчани износ (семантички маркер <money.approx>)
неких 14 милиона долара (1)
 - мера – опсег новчаног износа (семантички маркер <money.range>)
25 до 55 пенија (1), 5 до 25 USD (1), 80 и 100 милиона фунти (1)
 - мера – тачно одређена количина нечега (семантички маркер <amount.exact>)
три базе података (14), 10 бодова (13), две речи (10)
 - мера – приближно одређена количина нечега (семантички маркер <amount.approx>)
неколико десетина хиљада картона (2), на хиљаде свезака (2), неколико хиљада корисника (1)
 - мера – опсег количине нечега (семантички маркер <amount.range>)
два до три библиографска прилога (1), два до три пута (1), две до три научне области (1)
 - мера – тачно одређен проценат (семантички маркер <percent.exact>)
један проценат (2), десет посто (1), 95 посто (1)

- мера – опсег процента (семантички маркер <percent.range>)
1 до 5 процена (1)
- време – апсолутно одређен датум (семантички маркер <time.date.abs>)
1986. године (24), 1975. године (22), 2004. године (20)
- време – релативно одређен датум (семантички маркер <time.date.rel>)
данас (253), јануар (25), недеља (23)
- време – временски период (семантички маркер <time.date.period>)
1914-1918 (11), јануар-фебруар 1976 (9), новембар-децембар 1976 (8)
- време – апсолутно време у току дана (семантички маркер <time.hour.abs>)
10 сати увече (1), један поподне (1), три ујутру (1)
- време – релативно време у току дана (семантички маркер <time.hour.rel>)
зора (24), јутро (8), увече (4)
- време – период у току дана (семантички маркер <time.hour.period>)
9 до 22 сата (1)
- време – апсолутно трајање (семантички маркер <time.duration.abs>)
две године (21), пет година (21), три године (19)
- време – релативно трајање (семантички маркер <time.duration.rel>)
неколико година (20), последњих година (11), тих година (10)
- време – период трајања (семантички маркер <time.duration.period>)
60-их година (3), 80-их година (3), 4-5 година (2)
- време – понављање у времену (семантички маркер <time.set>)
свакодневно (23), годишње (23), сваке године (15).

Из наведених примера видимо да се у анализираном доменском корпусу јављају најразличитије поткатогеорије именованих ентитета. Није случајно што се тип организације *библиотека*, пуно име особе *Стојан Новаковић* или назив организације *IFLA Council* јавља са тако високом фреквенцијом – реч је о корпусу текстова из области библиотекарства и информатике.

Табела 9 – Фреквенција именованих ентитета према врстама у доменском корпусу

CategoryID	SubCategID	Caption in Serbian	Frequency
person	pers	O kome se radi?	6909
person	function	Čime se bavi?	528
organization	organization	O čemu se radi?	4047
location	location	Na kom mestu/gde?	4251
count	measure	Koliko?	464
count	money	Koliko novca?	46
count	amount	Koliko čega?	3308

count	percent	Koliko procenata?	5
time	date	Kog dana?	3120
time	hour	U koliko sati?	71
time	duration	Koliko to traje?	494
time	frequency	Koliko puta se ponavlja?	139

Целокупна листа именованих ентитета у доменском корпусу садржи 10.840 препознатих именованих ентитета из наведених категорија и поткатегорија. У табели 9 дат је преглед фреквенције именованих ентитета препознатих у доменском корпусу према врстама.

Размотримо посебно именоване ентитете који се односе на имена особа јер се они јављају са највећом фреквенцијом (у 6.909 случајева). Поред три најфреквентнија имена која су наведена у примерима, веома је интересантно да је доменски корпус богат извор информација и о другим знаменитим личностима српске културе. Међу препознатим именованим ентитетима пуних имена особа су и: *Георгије Магарашевић (25)*, *Павле Стефановић (22)*, *Лукијан Мушицки (21)*, *Иво Андрић (18)*, *Јован Суботић (15)*, *Урош Џонић (15)*, *Вук Караџић (13)*, *Ђура Даничић (12)*, *Милош Поповић (12)*, *Александар Поповић (12)*, *Захарије Орфелин (10)*, *Димитрије Давидовић (10)* и др.

Не изненађује да се од свих пуних имена особа највише пута јавља име *Светозар Марковић* пошто је име овог познатог српског књижевника и социјалисте саставни део назива Универзитетске библиотеке „Светозар Марковић” у Београду. Много аутора чланака из корпуса часописа „Инфотека” долази из ове установе, о овој библиотеци говори се у истом том корпусу, као и у монографији „Високошколске библиотеке.”

Као што видимо у табели 9, називи организација појављују се 4.047 пута. У листи препознатих именованих ентитета ове врсте потражили смо називе институција *Универзитетска библиотека „Светозар Марковић”* и *Народна библиотека Србије* јер смо у анализама екстрахованих вишечланих лема закључили да ће ови називи бити препознати као именовани ентитети. Утврдили смо да се међу именованим ентитетима назив прве институције јавља у више облика – *Универзитетска библиотека „Светозар Марковић” (32 пута)*, *Универзитетска библиотека у Београду (6)*, *УНИВЕРЗИТЕТСКА библиотека „Светозар Марковић” (6)*, *УНИВЕРЗИТЕТСКА БИБЛИОТЕКА „СВЕТОЗАР МАРКОВИЋ” (2)*, као и назив друге институције – *Народна библиотека Србије (38)*, *Народна библиотека Србије (НБС)*

(1), *НАРОДНА БИБЛИОТЕКА СРБИЈЕ (1)*. Сви наведени облици означени су семантичким маркером <org> и јављају се у различитим флективним облицима.

Листа препознатих локација (укупно 4251) доноси информације да се у доменском корпусу, осим података о *Београду*, *Новом Саду* и *Бечу*, могу пронаћи подаци и о *Крагујевцу (24)*, *Лондону (24)*, *Загребу (24)*, *Убу (24)*, *Паризу (22)*, *Сарајеву (20)*, *Вашиingtonу (19)*, *Бања Луци (18)* итд.

Остали именовани ентитети, осим апсолутно одређених датума, релативно одређених датума и временских периода, нису од нарочитог значаја за проналажење информација у овом доменском корпусу.

Закључићемо да је препознавање именованих ентитета, поред малог броја грешака, изузетно успешно извршено на анализираном корпусу текстова, на шта, осим броја препознатих ентитета, указује и њихова различитост. На основу свега наведеног можемо рећи и да ова технологија доприноси квалитетнијем проналажењу информација систематским издвајањем појмова и пружањем информације кориснику да ли је конкретан корпус релевантан за његово истраживање.

6.2.4 Анализа примене лексичких ресурса на новом тексту

Да бисмо верификовали допуњене речнике и проверили какве резултате они дају у екстракцији вишечланих термина из области библиотекарства и информатике приступили смо анализи на новом доменском корпусу. Нови доменски корпус чине текстови две докторске дисертације^{444, 445} из поменуте области одбрањене на Филолошком факултету Универзитета у Београду.⁴⁴⁶

За потребе анализе користили смо пречишћен текст докумената што подразумева да су изузети апстракти и кључне речи на енглеском језику, садржај, попис табела и прилога, табеле и слике у тексту, текст у фуснотама, као и регистри, прилози и библиографија у тексту и на крају дисертација јер, са једне стране, нису од значаја за екстракцију вишечланих термина из домена, а са друге стране, због конверзија PDF формата у чисти формат изгубила се континуација у оквиру редова у табелама и на

⁴⁴⁴ Драгана А. Грујић. „Школска библиотека у предметној и разредној настави основне школе” (Докторска дисертација. Филолошки факултет, 2013), у даљем тексту дисертација I

⁴⁴⁵ Александра М. Павловић. „Међубиблиотечка позајмица у Србији : развој и трансформација у складу са трансформацијом комуникационог система у науци” (Докторска дисертација. Филолошки факултет, 2009), у даљем тексту дисертација II

⁴⁴⁶ Љубазношћу ауторки достављен нам је коначни текст дисертација у електронском облику у PDF формату и дато одобрење за анализу на чему смо изузетно захвални.

сликама, па би претраживање резултирало са превише грешака. Такође, користили смо латиничну верзију текста јер смо комплетна претходна истраживања, екстракцију лема и исписе вишечланих речи вршили латиничним писмом. Врста писма није од значаја за истраживање ове врсте и промена писма не даје различите резултате.

Након поступка пречишћавања, ове текстове смо обрадили помоћу програма Unitex и добили следеће резултате: дисертација I садржи 1.687 реченица са 27.152 просте речи, а дисертација II садржи 1.566 реченица са 42.153 просте речи.⁴⁴⁷

У излиставању конкорданци и даљој анализи задали смо само критеријум приказивања вишечланих речи јер смо и допуну речника базирали на вишечланим терминима, а и зато што су те речи у нашем терминолошком речнику систематски обележене маркерима који су везани за домен. За просте речи оваква врста обележавања није систематски извршена, а додатни разлог што нисмо анализирали просте речи је и тај што су просте речи вишезначније од вишечланих речи.

Да бисмо упоредили резултате претраге пре и после допуне терминолошког речника, најпре смо анализу извршили помоћу речника који смо оформили на основу Библиотекарског терминолошког речника Народне библиотеке Србије. Тај речник, након сређивања класа речи, прочишћавања и одбацивања садржао је 2.323 вишечлане речи из домена библиотекарства и информатике и класа које су описане у одељку 4.4. Применом тог речника, добили смо резултате приказане у првом делу табеле 10 – у дисертацији I добили смо 99 погодака, односно конкорданци, а у дисертацији II 404 погодка, односно конкорданце.

Табела 10 – Вишечлани термини препознати у новом тексту

Препознати вишечлани термини	Дисертација I	Дисертација II
Првобитни речник	99	404
Допуна речника	474	1.036
Допуњени речник	562	1.358

У следећем кораку претрагу смо вршили само вишечланим речима из домена библиотекарства и информатике које смо одабрали за допуну терминолошког речника. Ова претрага је дала боље резултате: у дисертацији I добили смо 474 погодка, односно конкорданце, а у дисертацији II 1.036 погодака, односно конкорданце (други део табеле 10).

⁴⁴⁷ Аутоматско пребројавање реченица не мора да буде апсолутно тачно, али тачност сегментације и није од значаја за ово истраживање.

Када смо анализу извршили допуњеним речником вишечланих речи из домена библиотекарства и информатике, утврдили смо да је у дисертацији I пронађено 562 вишечлана термина, односно конкорданце, а у дисертацији II 1.358 вишечланих термина, односно конкорданци (трећи део табеле 10). Најфреквентнији термини који се јављају у овом новом корпусу су *међубиблиотечка позајмица* (294) и *школска библиотека* (156), затим *ауторско право* (104), *испука докумената* (72), *школски библиотекар* (40) итд. Овакав поредак термина по фреквенцији потпуно је очекиван имајући у виду теме о којима говоре анализирани докторске дисертације.

Уколико погледамо избор конкорданци из новог корпуса у прилозима 5а и 5б, видећемо да се јављају различити морфолошки облици истог вишечланог термина. Илустроваћемо ову тврдњу следећим примерима: у дисертацији I јављају се морфолошки облици *школска библиотека*, *школске библиотеке*, *школској библиотеци*, *школских библиотека*, *школским библиотекама*, *школску библиотеку* (прилог 5а), а у дисертацији II *међубиблиотечка позајмица*, *међубиблиотечке позајмице*, *међубиблиотечкој позајмици*, *међубиблиотечком позајмицом*, *међубиблиотечку позајмицу* (прилог 5б).

Анализираћемо још неке примере. У конкорданцама можемо приметити да су препознате вишечлане речи које су саставни део друге вишечлане речи. Погледаћемо следећи пример: *библотечко-информациона делатност* и *информациона делатност*. У конкорданцама налазимо примере за оба термина. Разлог за то је што су оба ова термина унета у термилошки речник. Следећи пример је: *васпитно-образовни процес* и *образовни процес*. Конкорданца за *васпитно-образовни процес* није означена јер се овај термин још увек не налази у термилошком речнику. У термилошком речнику постоји само термин *образовни процес* па је он препознат и пронађен (за оба примера видети прилог 5а).

Слична ситуација је у примеру за термине: *научне информације*, *техничке информације* и *научно-техничке информације*. Термини *научне информације* и *техничке информације* саставни су део термилошког речника, а термин *научно-техничке информације* не постоји у речнику. Зато је препознат само део овог термина *техничке информације* (видети прилог 5б).

Када у даљем раду, изван овог истраживања, речник буде допуњен и надограђен вишечланим терминима који су као нови екстраховани из овог корпуса, термини *васпитно-образовни процес* и *научно-техничке информације* биће уврштени у речник, па ће се наћи и у конкорданцама.

Закључићемо да, без обзира на то што је потребно радити на даљој надградњи, допуњени термилошки речник из библиотекарства и информатике који смо израдили за потребе овог истраживања даје боље резултате у проналажењу информација у односу на претходни електронски речник из поменуте области.

Табела 11 – Кандидати за термилошки речник са највећом фреквенцијом у новом корпусу

Disertacija I	FrekGraFLema	Disertacija II	FrekGraFLema
školski bibliotekar	40	međunarodna pozajmica	25
medijska pismenost	17	integracija elektronskih resursa	13
obrazovno-vaspitni rad	15	komunikacija u nauci	12
razredna nastava	14	sistem međubibliotečke pozajmice	12
čitalačke navike	10	naučni članak	10

Свакако је занимљиво да погледамо и који су се нови потенцијални термини јавили након аутоматске екстракције вишечланих речи из новог корпуса. У табели 11 навели смо по пет најфреквентнијих кандидата за термилошки речник из обе посматране дисертације.

Примена алата за препознавање именованих ентитета (NER) у новом корпусу

Као и у случају доменског корпуса и на новом корпусу смо применили алат за препознавање именованих ентитета. Укупан број препознатих именованих ентитета у новом корпусу износи 1.034 ентитета из свих категорија. Преглед фреквенције именованих ентитета према категоријама наведен је у табели 12.

Табела 12 – Фреквенција именованих ентитета према врстама у новом корпусу

CategoryID	SubCategID	Caption in Serbian	Frequency
person	pers	O kome se radi?	303
person	function	Čime se bavi?	21
organization	organization	O čemu se radi?	380
location	location	Na kom mestu/gde?	656
count	measure	Koliko?	11
count	money	Koliko novca?	2
count	amount	Koliko čega?	220
count	percent	Koliko procenata?	1
time	date	Kog dana?	302
time	hour	U koliko sati?	8
time	duration	Koliko to traje?	66
time	frequency	Koliko puta se ponavlja?	52

Прегледом листе препознатих именованих ентитета у два анализирана текста, утврдили смо да је међу именима особа поново најфреквентније име *Светозар Марковић* са 24 појављивања. Разлог за то је што се једна од дисертација у неким

деловима односи на анализу пословања Универзитетска библиотеке „Светозар Марковић” у домену међубиблиотечке позајмице. Навешћемо још неколико имена особа која бележе више појављивања: *Јован Јовановић*, *Стефан Немањић*, *Никола Тесла*. Од општих имена организација у овом корпусу текстова најчешће се, што је и логично, јавља *библиотека (22)*, а од назива институција *Универзитетска библиотека (11)*. Међу локацијама које означавају назив државе највећу фреквенцију има *Србија (160)*, а потом *Југославија (50)*, а од локација које се односе на насељено место најчешћи назив је *Београд (21)*.

Због величине новог копруса добили смо краћу листу препознатих именованих ентитета и не тако велику разноврсност у поткатогијама у односу на доменски корпус, али су и ови резултати веома значајни јер показују функционалност примене алата за препознавање именованих ентитета.

7 Закључак и план за даљи рад

7.1 Закључак

У оквиру ове дисертације бавили смо се истраживањем ефикасности и прецизности проналажења информација у дигиталним библиотекама, односно проналажењем дигиталних објеката који су опремљени дескрипторима метаподатака (субјективним дескрипторима) или дескрипторима садржаја (објективним дескрипторима), односно и једним и другим дескрипторима.

Први део истраживања је теоријске природе и односи се на преглед развоја библиотеке од традиционалне, преко електронске до дигиталне. Уз дефинисање свих наведених појмова, детаљно смо објаснили поступак увођења електронске библиотеке у нашу средину као прелазног облика између традиционалне и дигиталне библиотеке. Објаснили смо како проналажење информација функционише у традиционалним и електронским библиотекама. Потом смо анализирали проблеме пораста количине информација који су узроковали промене носилаца информација у савременом добу, али и проблем њихове застарелости. Дали смо и преглед особености дигиталних објеката у односу на аналогне изворе јер су те особине значајне за процес проналажења информација.

Други део овог истраживања посветили смо техникама за проналажење информација у дигиталним библиотекама. У оквиру тог дела најпре смо дефинисали дескрипторе метаподатака, представили различите врсте метаподатака (описне, административне, структуралне) и формате и схеме за њихов унос (XML, RDF, Даблинско језгро, MODS, METS, LOM, TEI). Затим смо у оквиру описа система за унос података о дигиталним објектима (Узајамна каталожко-библиографска база COBIB.SR, Дигитални репозиторијум Универзитета у Београду – PHAIDRA, Дигитална библиотека Филолошког факултета у Београду, Европеана) навели и анализирали начин уноса и врсте метаподатака које сваки од система садржи, те детаљно описали колекције домаћих библиотека садржаних на порталу Европеана које смо претраживали ради анализе проналажења информација (дигиталне колекције о Александру Великом, оријенталним рукописима, старим српским новинама, Првом светском рату). Указали смо и на формате за метаподатке развијене у оквиру пројеката Европеане (ESE, EDM) и на везу записа на порталу Европеана са семантичким вебom (Отворени повезани подаци – Linked Open Data).

Потом смо дефинисали дескрипторе садржаја и појмове ручно и аутоматско индексирање, одзив, прецизност и фреквенција појављивања термина и описали различите технологије за пречишћавање текста (оптичко препознавање карактера, оптичку сегментацију чланака, препознавање именованих ентитета) које доприносе квалитетнијем проналажењу информација коришћењем дескриптора садржаја. На крају другог дела истраживања анализирали смо лексичке ресурсе и алате који омогућавају боље проналажење информација у документима са комплетним текстом на српском језику (Корпус савременог српског језика, морфолошки електронски речник српског језика, српски wordnet, термилошки речник из библиотекарства и информатике, копрус поравнатих текстова из часописа „Инфотека” на српском и енглеском језику у оквиру лексичког алата *Библиша*).

Трећи део истраживања се у првом реду бави претрагом записа у наведеним системима за унос метаподатака о објектима, односно у дигиталним библиотекама које нуде претрагу преко комплетног текста. Постављајући најразличитије упите и претражујући записе за дигиталне објекте, као и садржај самих дигиталних објеката утврдили смо да, без обзира на све функционалности које нуде различити формати за унос метаподатака, као и везе ка семантичком вебу, претрага преко метаподатака не даје добре резултате у смислу прецизности. Одзив је углавном висок, међутим број релевантних докумената је углавном изузетно низак. Иако су у оквиру пројеката за потребе анализираних система развијани нови формати за унос метаподатака који су претпостављали већу ефикасност у проналажењу информација и везе ка другим изворима на мрежи, претрага ипак није резултирала квалитетним и прецизним проналажењем информација. С друге стране, претрага преко комплетног текста коришћењем дескриптора садржаја нуди неупоредиво боље резултате нарочито у оним дигиталним колекцијама које су опремљене квалитетним лексичким ресурсима, у првом реду морфолошким речницима.

У складу с тим је и четврти део истраживања у коме је фокус стављен на надградњу и допуну постојећих термилошких лексичких ресурса у циљу што бољег проналажења информација из области библиотекарства и информатике. Стога је у првом делу дат теоријски оквир, дефинисан појам терминологија и описана терминологија као наука, као и стандардизација у овој области (ISO стандарди, стандарди IFLA-е и српски стандарди који се тичу терминологије), а затим наведене и специфичности терминологије из библиотекарства и информатике и описани изабрани класични термилошки речници и лексикони, као и поједини тезауруси (у

штампаном и електронском облику). Анализирани су и стандарди за израду термилошких база података и описане карактеристике и начин уноса термина у базу *Терми*.

Практични део урађен у оквиру четвртог дела овог истраживања односи се на допуну термилошког речника српског језика из области библиотекарства и информатике. Конкретно, у процесу допуне речника, извршена је анализа текстова у доменском корпусу, испитана функционалност коначних трансдуктора у екстракцији вишечланих лема из поменутог доменског корпуса, проверена је исправност препознатих лема које се у доменском корпусу јављају са фреквенцијом од 4 и више пута, урађена тростепена евалуација екстрахованих вишечланих речи (исправност вишечлане речи, исправност леме, одабир термина за речник), допуњен термилошки речник одабиром 592 нове вишечлане речи, надограђен општи речник српског језика за 129 нових вишечланих термина, као и још 94 вишечланих термина који припадају и термилошком и општем речнику српског језика, а започет је и унос термина из поменуте области у термилошку базу *Терми*.

Након наведених анализа и допуне речника, проверена је функционалност допуњеног речника на новом корпусу текстова која је резултирала одличном прецизношћу у проналажењу информација и бољем препознавању термина у односу на првобитни речник. Даље, на текстовима из доменског корпуса примењен је и алат за препознавање именованих ентитета и проверена ефикасност претраге помоћу тог алата. Претрага је дала одличне резултате у виду од око 10.000 препознатих ентитета из различитих категорија и поткатегија. Ова врста претраге извршена је и на новом корпусу и, такође, је дала одличне резултате.

Ако размотримо оно што је урађено и представљено у овом раду, можемо рећи да су испуњени постављени циљеви и кроз детаљну анализу и низ примера утврђено да напредне методе помажу у квалитетнијем проналажењу информација. То је потврђено и упоређивањем ефикасности и прецизности проналажења информација уз примену напредних метода у односу на класичне методе. Такође, можемо рећи да проналажење информација помоћу дескриптора садржаја даје немерљиво боље резултате у односу на проналажење информација помоћу дескриптора метаподатака.

Почетна хипотеза овог рада такође је потврђена јер је анализом претраге у разним системима и дигиталним библиотекама утврђено да примена напредних метода и технологија на корпусу дигиталних текстова омогућава боље и прецизније проналажење информација у односу на оне колекције дигиталних текстова које нису

опремљене таквим методама и технологијама. У првом реду реч је о лексичким ресурсима међу којима посебно место заузимају морфолошки речници што је од посебног значаја за српски језик. У изградњи нових дигиталних библиотека, репозиторијума, архива и музејских колекција треба имати у виду добре резултате који су постигнути применом напредних метода и технологија на анализираним корпусима дигиталних библиотека не би ли се дошло до високе прецизности у проналажењу информација као и до повећања видљивости српског језика на оба писма у свету дигиталних података.

Рад на овом истраживању резултирао је и допуном општег електронског речника српског језика, као и допуном терминолошког речника из области библиотекарства и информатике. Тиме су у великој мери остварени предвиђени научни резултати.

7.2 План за даљи рад

Како је израда и допуна електронских речника захтеван и обиман посао, тако завршетком истраживања публикованог у овом раду није завршен рад на изградњи терминолошког речника из библиотекарства и информатике. Изван овог истраживања најпре је планиран наставак рада на речнику у смислу евалуације вишечланих речи и вишечланих лема које у доменском корпусу имају фреквенцију од три и мање појављивања, те одабир термина за допуну терминолошког речника на основу тих лема.

Ради повећања прецизности и већег обима аутоматски екстрахованих лема, треба радити на корекцији и доради постојећих, као и изради нових трансдуктора за екстракцију термина, а потом и све те трансдукторе тестирати на доменском корпусу већег обима. Доменски корпус могао би да садржи и дигиталне текстове докторских дисертација, мастер радова и семинарских радова одбрањених на Катедри за библиотекарство и информатику Филолошког факултета Универзитета у Београду, као и дигитализоване текстове радова из библиотекарства на српском језику старије од 70 година који не подлежу закону о ауторском праву.

Пошто смо у истраживању констатовали неопходност постојања терминолошких база података које користе међународне стандарде ради преноса, размене података, њихове обраде и поновне употребе, потребно је наставити рад на уносу термина из библиотекарства и информатике у базу *Терми*. У овој бази ће за сваки термин на српском језику бити, за почетак, унет његов еквивалент на енглеском језику са дефиницијама појма и његовим надређеним и подређеним концептима. Овако уређена

терминолошка база ће омогућити и аутоматску израду једнојезичних, двојезичних и вишејезичних речника.

Све набројане активности које су планиране као наставак рада на овом истраживању имају за циљ да допринесу квалитетнијем проналажењу информација применом дескриптора садржаја, односно објективних дескриптора у дигиталним колекцијама које садрже најразличитије документе.

8 Библиографија

8.1 Коришћена литература

Adamič, Stefan (ur.). *Sistem naučnih i tehnoloških informacija Jugoslavije : studija izvodljivosti*.

Beograd: Savez inženjera i tehničara Jugoslavije, 1988.

Allen, Richard E. (ed). *The Pocket Oxford Dictionary of Current English*. Oxford: Clarendon Press, 1984.

Anderson, Margaret J. *Carl Linnaeus: Father of Classification*. Berkeley Heights, NJ: Enslow Publishers, 1997.

Anić, Vladimir i Ivo Goldstein. *Rječnik stranih reči i izraza*. Zagreb: Novi Liber, 2004.

Arms, William Y., Christophe Blanchi, Edward A. Overly. „An Architecture for Information in Digital Libraries.” *D-Lib Magazine*, 3, 2(1997),

<http://www.dlib.org/dlib/february97/cnri/02arms1.html> (приступљено 14.07.2014)

Arms, William Y. *Digital libraries*. Cambridge, Massachusetts, London: The MIT Press, 2000.

Auger, Pierre. „La terminotique et les industries de la langue.” *Meta: journal des traducteurs*, 34,3(1989):450-456, <https://www.erudit.org/revue/meta/1989/v34/n3/001922ar.pdf> (приступљено 27.11.2015)

Auger, Pierre. „La terminologie au Quebec et dans le monde, de la naissance à la maturité.” In: *Actes du sixième colloque OLF-STQ de terminologie. L'ère nouvelle de la terminologie*, pp. 27-59. Quebec: Gouvernement du Québec, 1988.

Barker, Philip. „Electronic libraries — visions of the future.” *The Electronic Library* 12, 4(1994):221-230.

Бегенишић, Добрила. „Стандардизација библиотечке терминологије уз примену нових технологија.” У: Вранеш, Александра и Марковић, Љиљана (ур.) *Србија између истока и запада : наука, образовање, култура, уметност : тематски зборник у 4 књиге. Књ. 4, Језици Балкана у компаративном и интердисциплинарном контексту*. Београд : Филолошки факултет Универзитета, 2014.

Berners-Lee, Tim. *Metadata Architecture* <http://www.w3.org/DesignIssues/Metadata.html> (приступљено 08.07.2014)

Berners-Lee, Tim. „Linked Data.” World Wide Web Consortium, 2006, <http://www.w3.org/DesignIssues/LinkedData.html> (приступљено 10.09.2014)

Berners-Lee, Tim, James Hendler, Ora Lassil. „The Semantic Web.” *Scientific American Magazine* (May 17, 2001)

Bibliša. Journal Selection, <http://hlt.rgf.bg.ac.rs/Biblisha/Documentation.aspx> (приступљено 07.12.2015)

Bizer, Christian, Tom Heath, Tim Berners-Lee. „Linked Data – the Story So Far.” *Linked Data, International Journal on Semantic Web and Information Systems*, special issue (2009), <http://tomheath.com/papers/bizer-heath-berners-lee-ijswis-linked-data.pdf> (приступљено 23.10.2015)

Borgman, Christine. „The invisible library.” In: Kresh, D. (ed.). *Whole Digital Library Handbook*. Chicago: American Library Association, 2007.

Брофи, Питер. „Мерење учинка за библиотеке 21. века.” *Гласник Народне библиотеке Србије* 6,1(2004):203-218.

Brown, Keith (ed.) *The Encyclopedia of Language and Linguistics*. Oxford: Elsevier, 2005.

Cabré, Teresa M.. *Terminology: Theory, Methods, and Applications*. (Amsterdam : John Benjamins Publishing Co, 1999.

Campbell, Lorna. „Learning Object Metadata (LOM).” In: S. Ross, M. Day (eds.) *DCC Digital Curation Manual*, <http://www.dcc.ac.uk/resource/curation-manual/chapters/learning-object-metadata> (приступљено 20.10.2015)

Campo, Ángela. „The Reception of Eugen Wüster’s Work and the Development of Terminology.” Dissertation. Université de Montréal, Faculté des études supérieures, 2012. https://papyrus.bib.umontreal.ca/xmlui/bitstream/handle/1866/9198/Campo_Angela_2012_these.pdf (приступљено 17.11.2015)

Caplan, Priscilla. *Metadata Fundamentals for All Librarians*. Chicago: American Library Association, 2003.

Chartron, Ghislaine, Brigitte Guyot, Thierry Lafouge, Sylvie Lainé-Cruzel, Geneviève Lallich-Boidin, Marie France Peyrelong et Jean-Michel Salaün. „Le document numérique: un objet fédérateur de recherche en sciences de l’information.” *Documentaliste – Sciences de l’information* 39, 6(2002), <http://www.adbs.fr/le-document-numerique-un-objet-federateur-de-recherche-en-sciences-de-l-information-15075.htm?RH=REVUE>(приступљено 27.06.2014)

COBISS.Net, <http://www.cobiss.net/default-sr.asp> (приступљено 19.10.2015)

COBISS.SR. Библиотеке чланице – листа свих библиотека, http://vbshome.vbs.rs/sr/o_cobissu/biblioteke-sc.asp (приступљено 09.11.2015)

COMARC/B format za bibliografske podatke, http://obrazovanje.vbs.rs/DesktopModulesCobissIZO/IzoManualtextLoader.aspx?manual=1_comarc_b_svn&id=-1 (приступљено 15.06.2014)

Courtois, Blandine, Max L. Silberztein. „Dictionnaires électroniques du français.” *Langue française* 87,1(1990):3-4.

Coyle, Karen. „Understanding Metadata and Its Purpose.” *Journal of Academic Librarianship*, 31, 2(2005):160-163.

Crow, Raym. *The Case for Institutional Repositories: A SPARC Position Paper*. Washington, DC: The Scholarly Publishing and Academic Resources Coalition, 2002, http://www.arl.org/sparc/bm~doc/ir_final_release_102.pdf (приступљено 19.08.2014)

Cyganiak, Richard and Anja Jentzsch. „Linking Open Data cloud diagram“, <http://lod-cloud.net> (приступљено 10.09.2014)

DCMI– Dublin Core Metadata Initiative <http://dublincore.org> (приступљено 17.8.2015)

Делјанин, Сандра. „Застаревање дигиталних информација.” *Инфотека* 13,1(2012):47-58

Дискусија о библиотекарској терминологији. *Библиотекар* 16,5/6(1964):305-315; 17,1/2(1965):4-8; 17,3/4(1965):113-117; 17,5/6(1965):318-324.

DOI – Digital Object Identifier, <http://www.doi.org/> (приступљено 30.06.2014)

- Dublin Core Metadata Initiative. „DCMI Metadata Terms”, <http://dublincore.org/documents/dcmi-type-vocabulary/index.shtml> (приступљено 21.10.2015)
- Dunning, Ted. „Accurate methods for the statistics of surprise and coincidence.” *Computation Linguistics* 19,1(1993):61-74.
- Eakins, John and Margaret Graham. *Content-based information retrieval*. Manchester: JISC Technology Application Programm, 1999.
- Erjavec, Tomaž. “MULTEXT-East: morphosyntactic resources for Central and Eastern European languages.” *Language Resources and Evaluation* 46, 1(March 2012):131-142.
- Ерјавец, Томаж. „Смернице иницијативе за кодирање текста и њихова локализација.” *Инфотека* 11, 1 (2010):3-15, http://infoteka.bg.ac.rs/pdf/Srp/2010-1/INFOTHECA_XI_1_April2010_3-15.pdf (приступљено 19.10.2015)
- Europeana. <http://www.pro.europeana.eu/web/guest/content> (приступљено 20.05.2014)
- Europeana. „Definition of the Europeana Data Model v5.2.5”, <http://pro.europeana.eu/documents/900548/0d0f6ec3-1905-4c4f-96c8-1d817c03123c> (приступљено 28.08.2014)
- Europeana. „Europeana Data Model Primer“, <http://pro.europeana.eu/documents/900548/770bdb58-c60e-4beb-a687-874639312ba5> (приступљено 28.08.2014)
- Europeana. „Europeana Semantic Elements Specification and Guidelines“, , <http://pro.europeana.eu/documents/900548/2eee7beb-b9d8-4532-a089-8e8d6df38ce7> (приступљено 28.08.2014)
- Feather, John and Paul Sturges (eds.) *International Encyclopedia of Library and Information Science*. London: Routledge, 2003.
- Филолошки факултет Универзитета у Београду. ”Дигитална библиотека Филолошког факултета – корисничко упутство.”, <http://digifil.fil.bg.ac.rs/hlp/cyrl/Uputstvo.pdf> (приступљено 1.3.2016)
- Frantzi, Katerina, Sophia Ananiadou and Hideki Mima. „Automatic recognition of multi-word terms: the C-value/NC-value method.” *International Journal on Digital Libraries* 3,2(2000):115-130.
- Garfield, Eugene. „A Tribute to Calvin N. Mooers, a Pioneer of Information Retrieval.” *The Scientist* 11, 4(March 17, 1997), <http://www.the-scientist.com/?articles.view/articleNo/18361/title/A-Tribute-To-Calvin-N--Mooers--A-Pioneer-Of-Information-Retrieval/> (приступљено 02.06.2014)
- Gelbukh, Alexander, Grigori Sidorov, Eduardo Lavin-Villa and Liliana Chanona-Hernandez, „Automatic Term Extraction Using Log-Likelihood Based Comparison with General Reference Corpus.” In C. Hopfe, Y. Rezgui, E. Métais, A. Preece & H. Li (eds.), *Natural Language Processing and Information Systems* , Vol. 6177, pp. 248-255. Berlin, Heidelberg: Springer, 2010.
- Глумац, Дивна. *Предметни каталог у библиотеци општег типа*. Београд: Народна библиотека Србије, 1980.
- Goffin, Roger. 1985. „La science terminologique.” *Terminologie et traduction*, 2(1985):9-29.
- Gorman, Majkl. *Naše neprolazne vrednosti: bibliotekarstvo u XXI veku*. Beograd: Filološki fakultet, 2007.

- Граовац, Јелена. „Класификација текста заснована на српском wordnet-у.” *Инфотека* 14,2(2013):2-17, http://infoteka.bg.ac.rs/pdf/Srp/2013-2/INFOTHECA_XIV_2_2013_2-17.pdf (приступљено 3.3.2016)
- Greenberg, Jane. „Metadata and World Wide Web.” In: *Encyclopedia of Library and Information Science*, 72, suppl. 35(2002)
- Grishman, Ralph and Beth Sundheim. „Message Understanding Conference 6: A Brief History”. In: COLING-1996, Copenhagen, Denmark, pp. 466-471, <http://www.aclweb.org/anthology/C96-1079> (приступљено 04.11.2015)
- Grubačić, Kosta. *Enciklopedijski leksikon bibliotekarstva*. Sarajevo: Zavod za izdavanje udžbenika, 1964.
- Gruber, Tom. Ontology. In: Ling Liu and M. Tamer Özsu (eds.) *Encyclopedia of Database Systems* (Heidelberg etc.: Springer-Verlag, 2009) <http://tomgruber.org/writing/ontology-definition-2007.htm> (приступљено 22.10.2015)
- Habib, Michael. *Toward Academic Library 2.0: Development and Application of a Library 2.0 Methodology*. A Master’s Paper for the M.S. in L.S degree, School of Information and Library Science of the University of North Carolina at Chapel Hill. Chapel Hill, N.C., USA, 2006.
- Хоџ, Гејл. „Метаподаци на лак начин“. *Гласник Народне библиотеке Србије* 1(2004):157-180
- IDF– International DOI Foundation, http://www.doi.org/doi_handbook/7_IDF.html (приступљено 30.06.2014)
- IEEE. *Standard for Learning Object Metadata*, <https://standards.ieee.org/findstds/standard/1484.12.1-2002.html> (приступљено 20.10.2015)
- IFLA. *Međunarodni standardni bibliografski opis (ISBD) – objedinjeno izdanje*. Beograd: Narodna biblioteka Srbije, 2009, http://www.ifla.org/files/assets/cataloguing/isbd/isbd-cons_2007-sr.pdf (приступљено 15.06.2014)
- IFLA. *Smernice za audiovizuelnu građu u bibliotekama i drugim institucijama*. Hag : IFLA, 2004. <http://archive.ifla.org/VII/s35/pubs/avm-guidelines04-sr.pdf> (приступљено 27.06.2014)
- IFLA Classification and Indexing Section. *Guidelines of Multilingual Thesauri*, <http://archive.ifla.org/VII/s29/pubs/Profrep115.pdf> (приступљено 19.1.2016)
- IFLA Classification and Indexing Section. *Multilingual Dictionary of Cataloguing Terms and Concepts (MulDiCat)*, <http://www.ifla.org/publications/multilingual-dictionary-of-cataloguing-terms-and-concepts-muldicat?og=70> (приступљено 19.1.2016)
- Ikonomov, Nikola and Milena Dobрева. „The making of... digital book.” *NCD Review* 13(2008):1-8, <http://elib.mi.sanu.ac.rs/files/journals/ncd/13/ncd13001.pdf> (приступљено 11.09.2014)
- Институт за стандардизацију Србије. *Стандарди комисије А037*, http://www.iss.rs/rs/standard/index.php?national_committee_id=552 (приступљено 19.1.2016)
- Ињац, Весна. „Народна библиотека Србије – пуноправан партнер у пројекту 'Европска библиотека'.” *Преглед НЦД* 7(2005):49-54.
- International DOI Foundation. „DOI – Digital Object Identifier. Data model“, https://www.doi.org/doi_handbook/4_Data_Model.html (приступљено 15.10.2015)
- International DOI Foundation. „DOI – Digital Object Identifier. Numbering“, http://www.doi.org/doi_handbook/2_Numbering.html (приступљено 30.06.2014)

- International DOI Foundation. „DOI – Digital Object Identifier. Operating procedures“, http://www.doi.org/doi_handbook/9_OperatingProcedures.html (приступљено 15.10.2015)
- International DOI Foundation. „DOI Handbook, History“, https://www.doi.org/doi_handbook/1_Introduction.html (приступљено 14.10.2015)
- ISBN agencija. *ISBN priručnik: međunarodno izdanje*. London: Međunarodna ISBN agencija, 2012.
- ISO – Interantional Standard Organization, <http://www.iso.org/iso/home/standards.htm> (приступљено 30.06.2014) ISO. Computer Applications in Terminology – Terminological Markup Framework, 2003. Ref. ISO16642:2003
- ISO. Computer Applications in Terminology – Machine-Readable Terminology Interchange Format (MARTIF) – Negotiated Interchange, 1999. Ref. ISO 12200:1999
- ISO 2788:1986 – Guidelines for the establishment and development of monolingual thesauri, http://www.iso.org/iso/catalogue_detail.htm?csnumber=7776 (приступљено 30.11.2015)
- ISO. Standards catalogue– 01.020: Terminology (principles and coordination), including terminography, http://www.iso.org/iso/iso_catalogue/catalogue_ics/catalogue_ics_browse.htm?ICS1=1&ICS2=20 (приступљено 27.11.2015)
- ISO. Systems to Manage Terminology, Knowledge and Content – TermBase eXchange (TBX), 2008. Ref. ISO 30042:2008
- ISO. Terminology work – Principles and methods, 2009. Ref. ISO 704:2009, http://www.iso.org/iso/catalogue_detail.htm?csnumber=38109 (приступљено 27.11.2015)
- Jacquet-Pfau, Christine. „Les dictionnaires du français sur cédérom.“ *International Journal of Lexicography* 15,1(2002):89-104.
- Jokanović, Vladimir, Emil Popović i Miloš Stojanović. *Mala jugoslovenska enciklopedija bibliotekarstva*. Beograd: Nova knjiga, 1988.
- Jokanović, Vladimir, Fahrudin Kalender, Emil Popović i Miloš Stojanović. *Bibliotekarski leksikon*. Beograd: Nolit, 1984.
- Kallinikos, Jannis, Aleksu Aaltonen and Atila Marton. „A theory of digital objects.“ *First Monday* 15, 6-7, (June 2010), <http://firstmonday.org/ojs/index.php/fm/article/view/3033/2564> (приступљено 27.06.2014)
- Конзорцијум библиотека Србије за обједињену набавку– КОBSON, http://www.kobson.nb.rs/kobson/o_konzorcijumu.1.html (приступљено 16.06.2014)
- Kosanović, Biljana. „DOI u Srbiji“, <http://www.doiserbia.nb.rs/img/DOI2010.pdf> (приступљено 15.10.2014)
- Kosanović, Biljana. „Naučne informacije u Srbiji – ponuda iskustva KOBSON-a.“ *Kultura* 129(2010):106-114, <http://scindeks-clanci.ceon.rs/data/pdf/0023-5164/2010/0023-51641029106K.pdf> (приступљено 30.06.2014)
- Ковачевић, Љиљана, Весна Ињац и Добрила Бегенишић. *Библиотекарски термиолошки речник*. Београд: Народна библиотека Србије, 2004.
- Ковачевић, Љиљана, Добрила Бегенишић, Весна Ињац-Малбаша. *Речник библиотекарства и информационих наука: српско-енглеско-немачки*. Београд: Народна библиотека Србије, 2014, <http://rbi.nb.rs/sr/home.html> (приступљено 30.10.2015)

- Kovačević, Ljiljana. „Srpska bibliotekarska terminologija u međunarodnom projektu: višejezički rečnik bibliotekarske terminologije.” *Glasnik Narodne biblioteke Srbije* 1(2000), https://www.nb.rs/view_file.php?file_id=512 (приступљено 30.11.2015)
- Krstev, Cvetana, Duško Vitas, „An Aligned English-Serbian Corpus”, In: N. Tomović & J. Vujić (eds.) *ELLSIIR Proceedings (English Language and Literature Studies: Image, Identity, Reality)*, Vol. 1, Belgrade, 4-6 December 2009, pp. 495-508. Belgrade: University, Faculty of Philology, 2011.
- Krstev, Cvetana, Ranka Stanković, Ivan Obradović, Duško Vitas and Miloš Utvić. „Automatic Construction of a Morphological Dictionary of Multi-Word Units.” In: Hrafn Loftsson, Eiríkur Rögnvaldsson, Sigrún Helgadóttir (eds.) *Proceedings of the 7th International Conference on NLP, IceTAL 2010*, Reykjavik, Iceland, August 16-18, 2010, pp. 226-237, Lecture Notes in Computer Science 6233. Heidelberg: Springer 2010.
- Krstev, Cvetana & Duško Vitas. „Corpus and Lexicon — Mutual Incompleteness.” In: P. Danielsson & M. Wagenmakers (eds), *Proceedings of the Corpus Linguistics Conference*, 14–17 July 2005, Birmingham, , <http://www.birmingham.ac.uk/research/activity/corpus/publications/conference-archives/2005-conf-e-journal.aspx> (приступљено 12.10.2015)
- Крстев, Цветана. „Дигиталне библиотеке – разграничење појмова.” *Инфотека* 3, 1-2(2002):3-14.
- Krstev, Cvetana. „Digitalne biblioteke kao potencijalni lingvistički resurs – stanje u Srbiji.” In: Voss, Ch. und Golubović, B. (hrsg.) *Srpska lingvistika– Serbische Linguistik*, SLCCEE, Bd. 7, pp 123-142. München, Berlin: Otto Sagner, 2010.
- Krstev, Cvetana i Duško Vitas. „Informatički pogled na bibliografiju.” U: Aleksandra Vraneš (ur.), *Srpska bibliografija danas*, str. 229-241. Novi Sad: Matica srpska, 2008.
- Крстев, Цветана, Бојана Ђорђевић, Сања Антонић, Невена Ивковић-Берчек, Зорица Зорица, Весна Црногорац, Љиљана Мацура. „Кооперативни рад на доградњи српског Wordnetа.” *Инфотека* 9,1-2(2008): 57-75.
- Krstev, Cvetana. „Kurs iz XML-a.”, <http://poincare.matf.bg.ac.rs/~cvetana/kurs-xml/xml-codes/xmlcodes8.html> (приступљено 15.10.2015)
- Krstev, Cvetana. *Materijali za kurs Napredne metode u pronalaženju informacija*, nepublikovani materijal, , <http://poincare.matf.bg.ac.rs/~cvetana/Nastava/Materijal/NM-PI-03.ppt> (приступљено 20.10.2015)
- Krstev, Cvetana. *Materijali za kurs Pronalaženje informacija, deo Automatsko indeksiranje*, nepublikovani materijal za školsku 2014/2015.
- Krstev, Cvetana. *Materijali za kurs Pronalaženje informacija, deo Automatsko indeksiranje*, nepublikovani materijal za školsku 2015/2016.
- Krstev, Cvetana. *Processing of Serbian: Automata, Texts and Electronic Dictionaries*. Belgrade: Faculty of Philology, 2008.
- Krstev, Cvetana. „Specifični koncepti Balkana u semantičkoj mreži Wordnet.” U: Ljiljana Subotić et al (ur.) *Zbornik radova „Susreti kultura“*, Novi Sad, decembar 2004, pp. 275-285. Novi Sad: Univerzitet, Filozofski fakultet, 2006.
- Krstev, Cvetana, Ivan Obradović, Miloš Utvić and Duško Vitas. „A System for Named Entity Recognition Based on Local Grammars.” *Journal of Logic and Computation* 24, 2(2014):473-489.

- Krstev, Cvetana, Duško Vitas and Gordana Pavlović-Lažetić. „Resources and methods in the morphosyntactic processing of Serbo-Croatian.” In Gerhild Zybatow et al. (eds.) *Formal Description of Slavic Languages: The Fifth Conference*, Leipzig 2003, pp. 3-17. Frankfurt am Main, Peter Lang 2008.
- Krstev, Cvetana. „Specifični koncepti Balkana u semantičkoj mreži Wordnet.” U: Ljiljana Subotić et al (ur.) *Zbornik radova „Susreti kultura“*, Novi Sad, decembar 2004, pp. 275-285. Novi Sad: Univerzitet, Filozofski fakultet, 2006.
- Krstev, Cvetana, Duško Vitas, Vesna Šatev. “TEI Encoding of Serbian Proverbs.” In: Anisava Miltenova et al. (eds.) *Computer Applications in Slavic Studies*, Proceedings of Azbuky.Net International Conference and Workshop, 24-27 October 2005, Sofia, Bulgaria, pp. 101-114. Sofia: Institute of Literature, Bulgarian Academy of Sciences, 2006.
- Krstev, Cvetana, Ranka Stanković, Ivan Obradović and Biljana Lazić. „Terminology Acquisition and Description Using Lexical Resources and Local Grammars.” In: Thierry Poibeau and Pamela Faber (eds.) *Proceedings of the 11th Conference on Terminology and Artificial Intelligence, LexiCon*, Universidad de Granada, Granada, Spain, 2015, pp. 81-89, CEUR Workshop Proceedings, http://ceur-ws.org/Vol-1495/paper_13.pdf (приступљено 20.1.2016)
- Krstev, Cvetana, Svetla Koeva and Duško Vitas. “Towards the Global Wordnet.” In: *Conference Abstracts of the First International Conference of Digital Humanities Organizations (ADHO) Digital Humanities 2006*, Paris-Sorbonne, 5-9 July 2006, p. 114-117. Paris: Université Paris-Sorbonne, 2006.
- Krstev, Cvetana, Ranka Stanković, Duško Vitas, Ivan Obradović. „The Usage of Various Lexical Resources and Tools to Improve the Performance of Web Search Engines.” In: *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08)*, Marrakech, Morocco, 28-30 May 2008, pp. 219-224. European Language Resources Association (ELRA), 2008, <http://www.lrec-conf.org/proceedings/lrec2008/> (приступљено 30.10.2015)
- Krstev, Cvetana, Gordana Pavlović-Lažetić and Ivan Obradović. „Using textual and lexical resources in developing Serbian Wordnet.” *Romanian Journal of Information Science and Technology* 7,1-2(2004):147-161.
- Krstev, Cvetana, Ranka Stanković, Duško Vitas, Ivan Obradović, „WS4LR– a Workstation for Lexical Resources”. In: *Proceedings of the Fifth International Conference on Language Resources and Evaluation*, Genoa, Italy, May 2006, pp. 1692-1697, 2006, <http://www.lrec-conf.org/proceedings/lrec2006/> (приступљено 30.10.2015)
- Lexical Computing, „Statistics used in Sketch Engine”, <https://www.sketchengine.co.uk/wp-content/uploads/ske-stat.pdf> (приступљено 17.2.2016)
- Luhn, Hans Peter. „The automatic creation of literature abstracts.” *IBM Journal of Research and Development* 2(1958):159-165.
- Lynch, Clifford A. „Institutional Repositories: Essential Infrastructure for Scholarship in the Digital Age.” *ARL* 226(2003):1-15, <http://www.arl.org/bm~doc/br226ir.pdf> (приступљено 19.08.2014)
- Machine Readable Catalog – MARC21 – format for bibliographical data, <http://www.loc.gov/marc/bibliographic/ecbdhome.html> (приступљено 15.06.2014)
- Macmillan Dictionary*. Nomenclature, <http://www.macmillandictionary.com/dictionary/british/nomenclature> (приступљено 17.11.2015)

- Macmillan Dictionary. Terminology*, <http://www.macmillandictionary.com/dictionary/british/terminology> (приступљено 18.11.2015)
- Мала енциклопедија Просвета: општа енциклопедија*. Београд: Просвета, 1959.
- Манцуков, Митко и Јасмина Рацковић. „Примена аутоматске обраде података у Библиотеци града Београда.” *Библиотекар* 40,3/4(1988):15-32.
- MarcLogic. „Language Support“, http://www.marklogic.com/resources/marklogic-language-support/resource_download/datasheets/ (приступљено 21.12.2015)
- McEnery, Anthony, Richard Xiao and Yukio Tono. *Corpus Based Language Studies: an Advance Resource Book*. London: Routledge, 2006.
- Meghini, Carlo, Nicolas Spyrtos, Jitao Yang. „A data model for digital libraries.” *International Journal on Digital Libraries* 11(2010):41-56.
- Melby, Alan K. „Terminology in the Age of Multilingual Corpora.” *The Journal of Specialized Translation* 18(2012):7-29, http://www.ttt.org/tbx/art_melby-JoSTransNum18.pdf (приступљено 11.12.2015)
- Merriam-Webster Dictionary. Descriptor*, <http://www.merriam-webster.com/dictionary/descriptor> (приступљено 7.7.2014)
- META-SHARE. Multilingual Edition of Verne’s Novel “Around the World in 80 Days”, <http://metashare.nlp.ipipan.waw.pl/metashare/repository/browse/multilingual-edition-of-vernese-novel-around-the-world-in-80-days/438d369a8bdf11e2b44a00151714459236eaa2be46a54847812a1f210d9f378e/> (приступљено 30.10.2015)
- МФЦ Микрокомерц. Софтвер Spider.Net дигитална библиотека – LibIT, <http://www.earhiva.com/softver/mfc/libit.php#close> (приступљено 25.11.2014)
- МФЦ Микрокомерц. Софтвер Spider.Net репозиторијум докумената, <http://www.earhiva.com/softver/mfc/repozitorijum.php> (приступљено 25.11.2014)
- Михайлов, А. И. (ред.). *Терминологическиј словарь по научной информации : (на руском, сербохрватском, словенском и македонском језици) = Terminološki rečnik iz oblasti naučnih informacija : (na ruskom, srpskohrvatskom, slovenačkom i makedonskom jeziku) = Terminološki slovar iz područja znanstvene informacije : (v ruskem, srbohrvatskem, slovenskem i macedonskem jeziku) = Терминолошки речник за научна информација : (на руски, српско-хрватски, словеначки и македонски јазик)*. Београд: Institut za naučno-tehničku dokumentaciju i informacije, 1969.
- Milenković, Marko. „Inicijativa 'Dublin Core'.” *Pregled Nacionalnog centra za digitalizaciju*, 2(2003):70-79.
- Мишић, Смиља. „Израда српскохрватског библиотекарског речника: терминолошки проблеми.” *Библиотекар* 13,4(1961):227-241.
- Mkadmi, Abderrazak et Imad Saleh. *Bibliothèque numérique et recherche d'informations*. Paris: Lavoisier, 2008.
- Moens, Marie-Francine. *Automatic Indexing and Abstracting of Document Texts*. Dordrecht: Kluwer Academic Publishers, 2000.

- Mortimer, Mary. *Library Speak: a Glossary of Terms in Librarianship and Information Management*. Friendswood, USA: Total Recall Publications, 2007.
- Neavill, Gordon. „Emergence of Digital Libraries.” *Infotheca* 5,1-2(2004):25-33
- Neudecker, Clemens, Hans-Jörg Lieder and Sandra Cobel. *Europeana Newspapers – Final report*, http://www.europeana-newspapers.eu/wp-content/uploads/2015/05/Final_Report.pdf (приступљено 27.2.2016)
- Neudecker, Clemens and Lotte Wilms. Europeana Newspapers – a Gateway to Europeana Newspapers Online. In: *IFLA Newspapers/GENLOC Pre-Conference Satellite Meeting Newspapers to the People*, Singapore, 14 – 15 August 2013, http://www.ifla.org/files/assets/newspapers/Singapore_2013_papers/day_1_10_2013_ifla_satellite_n_eudecker_clemens_wilms_lotte_europeana_newspapers_-_a_gateway_to_european_newspapers_online.pdf (приступљено 26.08.2014)
- OCLC. WorldCat – a Global Library Resource, <http://www.oclc.org/worldcat/catalog.en.html> (приступљено 09.11.2015)
- Oeser, Erhard (Hrsg.). *Eugen Wüster (1898– 1977): Leben und Werk, ein österreichischer Pionier der Informationsgesellschaft*. Vienna : TermNet, 1998.
- Online Dictionary of Language Terminology (ODLT)*. Terminology, <http://odlt.org/> (приступљено 18.11.2015)
- Oxford English Dictionary– OED, online edition <http://www.oed.com/viewdictionaryentry/Entry/95568#eid417511> (приступљено 02.06.2014)
- Paskin, Norman. „Digital Object Identifier for scientific data.“ *Data Science Journal* 4, 1(18 March 2005):1-9, <http://www.doi.org/topics/041110CODATAarticleDOI.pdf> (приступљено 30.06.2014)
- Paumier, Sébastien. *Unitex 3.1 beta : user manual*. Paris : Université Paris-Est Marne-la-Vallée, 2003-2014 , <http://www-igm.univ-mlv.fr/~unitex/UnitexManual3.1.pdf> (приступљено 25.09.2014)
- Petit Robert*. Paris: Robert, 1978.
- Петровић, Јасмина. „Библиотека града Београда.” *Гласник Народне библиотеке Србије*, 4,1(2002):185-202, https://www.nb.rs/view_file.php?file_id=609 (приступљено 09.12.2015)
- Петровић, Петар М. (ур.). *Свезнање: општи енциклопедиски лексикон*. Београд: Народно дело, 1937.
- Радовић-Тешић, Милица. *Граматички и лингвистички појмовник*. Београд:Учитељски факултет, 2011.
- Ray, Alain. *Essays on Terminology*. Amsterdam : John Benjamins Publishing Co. 1995.
- Reitz, Joan M.. ODLIS – Online Dictionary for Library and Information Science, <http://www.abc-clio.com/ODLIS/searchODLIS.aspx> (приступљено 09.11.2015)
- Република Србија, Министарство културе и информисања. „Закон о обавезном примерку публикација, члан 4, став 3.” *Службени гласник Републике Србије*, 52(2011)
- Република Србија. „Закон о ауторском и сродним правима: чл. 103, 104 и 105.” *Службени гласник Републике Србије*, 104(2009), 99(2011), 119(2012)
- Ристовић, Зоран. „Од корпуса до учионице – примена паралелизованих текстова у настави енглеског језика у основној школи.” *Инфотека* 13, 2(2012):52-66.

- Salton, Gerard. *Automatic Text Processing– The Transformation, Analysis, and Retrieval of Information By Computer*. (Boston: Addison-Wesley Publishing Company, 1989)
- Schmitt, Norbert. *Vocabulary in language teaching*. Cambridge: Cambridge University Press, 2000.
- Sekine, Satoshi, Kiyoshi Sudo and Chikashi Nobata. „Extended Named Entity Hierarchy.” In: *Proceedings of the Third International Conference on Language Resources and Evaluation (LREC '02)*, Las Palmas, Canary Islands, Spain, May 29-31 2002, pp. 1818-1824, <http://www.lrec-conf.org/proceedings/lrec2002/pdf/120.pdf> (приступљено 04.11.2015)
- Sekine, Satoshi and Chikashi Nobata. „Definition, dictionaries and tagger for Extended Named Entity Hierarchy.” In: *Proceedings of the Forth International Conference on Language Resources and Evaluation (LREC '04)*, Lisbon, Portugal, May 26-28 2004, pp. 1977-1980, <http://www.lrec-conf.org/proceedings/lrec2004/> (приступљено 04.11.2015)
- Seljak, Tomaž i Tvrtno Šercar. „Prikaz razvoja Sistema naučnih i tehnoloških informacija Jugoslavije (SNTIJ).” U: *Zbornik referata III jugoslovenskog savetovanja Specijalizovani informacioni sistemi na II naučnom skupu o Sistemu naučnih i tehnoloških informacija*, Dubrovnik, 6-8. jun 1990, str. 9-18. Beograd: Savez inženjera i tehničara Jugoslavije, 1990.
- SETimes. O nama, http://www.setimes.com/cocoon/setimes/xhtml/sr_Latn/document/setimes/footer/about/about (приступљено 22.10.2014)
- Sofronijević, Adam and Jelena Mitrović. „Europeana Libraries Project: Fostering Digitization Across the Europe.” *NCD Review* 21(2012):57-64.
- Софронијевић, Адам. „Нова парадигма сарадње у библиотекама.“ Докторска дисертација, Филолошки факултет, 2015.
- Sowa, John F. Semantic Networks. In: Stuart C. Shapiro (ed.) *Encyclopedia of Artificial Intelligence*. Hoboken, NJ, Wiley, 1992. <http://www.jfsowa.com/pubs/semnet.pdf> (приступљено 14.10.2015)
- Stanković, Ranka, Ivan Obradović and Aleksandra Trtovac. „An Approach to Development of Bilingual Lexical Resources.” In: Zoran Budimac, Mirjana Ivanović and Miloš Radovanović (eds.) *Local Proceedings of the Fifth Balkan Conference in Informatics, 2012, BCI '12*, Novi Sad, Serbia, September 16-20, 2012, <http://ceur-ws.org/Vol-920> (приступљено 23.10.2014)
- Stanković, Ranka, Cvetana Krstev, Biljana Lazić and Dalibor Vorkapić. „A bilingual digital library for academic and entrepreneurial knowledge management.” In: J.C. Spender, Giovanni Schiuma and Vito Albino (eds.) *Proceeding of 10th International Forum on Knowledge Asset Dynamics — IFKAD 2015: Culture, Innovation and Entrepreneurship: connecting the knowledge dots*, Bari, Italy, 10-12 June 2015:1764-1777.
- Stanković, Ranka, Ivan Obradović and Miloš Utvić. „Developing Termbases under the TBX Standard.” In: Pavlović-Lažetić, Gordana et al. (eds.), *Natural Language Processing for Serbian*, pp. 12-26. Belgrade: University of Belgrade, Faculty of Mathematics, 2014.
- Stanković, Ranka, Ivan Obradović, Cvetana Krstev and Duško Vitas. „Production of morphological dictionaries of multi-word units using a multipurpose tool”, In: K. Jassem, P. W. Fuglewicz, M. Piasecki and A. Przepiórkowski (eds.) *Proceedings of the Computational Linguistics-Applications Conference*, October 17–19, 2011. Jachranka, Poland, pp. 77-84. Warsaw: Polish Information Processing Society, 2011.

- Stanković, Ranka, Cvetana Krstev, Ivan Obradović, Biljana Lazić and Aleksandra Trtovac. „Rule based automatic multi-word term extraction and lemmatization.” In: *10th edition of the Language Resources and Evaluation Conference (LREC)*, 23-28 May 2016, Portorož (Slovenia) (у штампи)
- Stanković, Ranka, Cvetana Krstev, Ivan Obradović, Aleksandra Trtovac and Miloš Utvić. „A tool for enhanced search of multilingual digital libraries of e-journals.” In: Nicoletta Calzolari et al. (eds.) *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12)*, Istanbul, May 23-25, 2012, pp. 1710-1717 (Istanbul: European Language Resources Association (ELRA), 2012), http://www.lrec-conf.org/proceedings/lrec2012/pdf/375_Paper.pdf (приступљено 07.12.2015)
- Стратегија даљег развоја информатике у Савезној Републици Југославији (пројекат за Савезну владу)*. Београд: Савезно министарство за развој, науку и животну средину, 1997.
- TEI – Text Encoding Initiative. History, <http://www.tei-c.org/About/history.xml> (приступљено 19.10.2015)
- TEI Consortium. TEI P5 – Guidelines for Electronic Text Encoding and Interchange, <http://www.tei-c.org/Vault/P5/2.9.1/doc/tei-p5-doc/en/Guidelines.pdf> (приступљено 19.10.2015)
- Tennant, Roy. „Column Digital Libraries: 21st-Century Cataloguing.” *Library Journal* (April 15, 1998), <http://roytennant.com/column/?fetch=data/96.xml> (приступљено 09.07.2014)
- Thompson, Antony. *Rječnik bibliotekarskih stručnih izraza: englesko-hrvatskosrpski*. Zagreb: Školska knjiga, 1965.
- Trtovac, Aleksandra and Jelena Andonovski. „Enrichment of Morphological Dictionary of MWUs.” In: Pavlović-Lažetić, Gordana et al. (eds.), *Natural Language Processing for Serbian*, pp. 27-40. Belgrade: University of Belgrade, Faculty of Mathematics, 2014.
- Тртовац, Александра. „Метаподаци у систему PHAIDRA – дигиталном репозиторијуму Универзитета у Београду.” У: Весна Црногорац, Весна Ињац (ур.) *Хоризонти светског и европског библиотекарства у дигиталном добу*, Међународна научна конференција, Београд 27-28. октобар 2011, стр. 171-188. Београд: Библиотекарско друштво Србије, 2012, <http://bds.rs/wordpress/wp-content/uploads/2013/08/Horizonti-svetskog-i-evropskog-bibliotekarstva-u-digitalnom-dobu.pdf> (приступљено 20.10.2015)
- Universal MARC format, Unimarc bibliographic format, <http://www.ifla.org/publications/unimarc-formats-and-related-documentation> (приступљено 15.06.2014)
- University of Vienna. *Phaidra: Permanent Hosting, Archiving and Indexing of Digital Resources and Assets*, http://www.westbulnet.com/AmplioCMS2/public/EnterPublic.cms2?langId=74&request_locale=sr_CS#menuId=1698 (приступљено 20.08.2014)
- Универзитет у Београду. *Е-тезе Универзитета у Београду*, <http://eteze.bg.ac.rs/> (приступљено 30.10.2015)
- Универзитетска библиотека „Светозар Марковић.” *Претраживе дигитализоване историјске новине*, <http://www.unilib.rs/istorijske-novine/pregled> (приступљено 18.2.2016)
- Utvić, Miloš. „Izgradnja referentnog korpusa savremenog srpskog jezika.” *Doktorska disertacija*, Filološki fakultet, 2013.
- Utvić, Miloš. „Konačni automati u regularnoj imenskoj derivaciji.” *Magistarska teza*, Matematički fakultet, 2008.

Утвић, Милош. „Листе учестаности Корпуса савременог српског језика.“ *Међународни славистички центар* 43,3(2014):241-262.

Утвић, Милош. „Регуларни изрази.“ *Свет речи* 15-16(2003):66-68.

Vitas, Duško and Cvetana Krstev. „Derivational Morphology in an E-Dictionary of Serbian.” In: Zygmunt Vetulani (ed.) *Proceedings of 2nd Language & Technology Conference*, April 21-23, 2005, Poznań, Poland, pp.139-143. Poznań: Wydawnictwo Poznańskie, 2005.

Вранеш, Александра. „Дигитална библиотека“, <http://www.fil.bg.ac.rs/lang/sr/biblioteke/digitalna-biblioteka/> (приступљено 20.08.2014)

Вранеш, Александра и Љиљана Марковић. *Од рукописа до библиотеке*. Београд: Филолошки факултет, 2008.

Вранеш, Александра. „Терминолошке двоумице у библиотекарству.” У: Младеновић, Добривоје (прир.), *Библиотекарство на крају века, Књ. 4*, стр. 65-68. Београд: Библиотекарско друштво Србије, 1999.

Вранеш, Александра. *Високошколске библиотеке*. Београд: Универзитетска библиотека „Светозар Марковић : Tempus пројекат, 2004.

Вујаклија, Милан. *Лексикон страних речи и израза*. Београд: Просвета, 1954.

Вујанић, Милица, Даринка Гортан-Премк, Милорад Дешић, Рајна Драгићевић, Мирослав Николић, Љиљана Ного, Васа Павковић, Никола Рамић, Рада Стијовић, Милица Тешић, Егон Фекете. *Речник српскога језика*. Нови Сад: Матица српска, 2007.

W3 Techs. „Usage of content languages for websites“, http://w3techs.com/technologies/overview/content_language/all (приступљено 14.10.2015)

Ware, Mark. *Pathfinder Research on Web-based Repositories*. London: Publisher and Library/Learning Solution, 2004, <http://mrkwr.files.wordpress.com/2006/11/pals-report-on-institutional-repositories.pdf> (приступљено 19.08.2014)

Webster's Third New International Dictionary of the English Language. Terminology. Unabridged ed. Springfield, Mass., G. & C. Merriam Co, 1966.

Wüster, Eugen. „L'étude scientifique générale de la terminologie, zone frontalière entre de la linguistique, la logique, l'ontologie, l'informatique et les science des choses.” In: H. Rondeau & G. Felber (eds.) *Textes choisis de terminologie. Vol. 1, Fondements théoretiques de la terminologie*, p. 55-113. Ville de Québec: Université Laval, GIRSTERM– Groupe interdisciplinaire de recherche scientifique et appliquée en terminologie, 1981.

XQuery 3.0 – an XML Query Language, <http://www.w3.org/TR/xquery-30/> (приступљено 7.12.2015)

8.2 Рефрентне веб локације

Adobe

<https://www.adobe.com/>

Adobe TIFF Resources Page

<http://partners.adobe.com/public/developer/tiff/index.html>

ALA – American Library Association. *Guidelines on Subject Access to Individual Works of Fiction, Drama (GSAFD)*

<http://experimental.worldcat.org/gsafd/>

Apple

<http://www.apple.com/>

BAEKTEL – Blending academic and entrepreneurial knowledge in technology enhanced learning

<http://www.baektel.eu/>

Basel Register of Thesauri, Ontologies and Classifications – BARTOC

<http://bartoc.org/en>

BBC – British Broadcasting Corporation

<http://www.bbc.com/>

Библиотека града Београда

<http://www.bgb.rs/>

Библиотекарски термилошки речник

<http://btr.nb.rs/default.htm>

Bibliothèque nationale de Luxembourg

<http://www.bnl.public.lu/fr/index.html>

Bibliothèque nationale de Luxembourg. Bnlviewer

<http://sourceforge.net/projects/bnlviewer/>

BISIS

<http://bisis.uns.ac.rs/wiki/WikiStart>

BNF, Data about BnF

<http://data.bnf.fr/>

CCER – Centre for Computer-aided Egyptological Research, *Multilingual Egyptian Thesaurus*

<http://projetrosette.info/page.php?Id=290>

CENL – Conference of European National Libraries

<http://www.cenl.org/>

COBIB.SR

<http://www.vbs.rs/scripts/cobiss?ukaz=INFO&id=1946024215347051&pg=1>

COBISS– Cooperative Online Bibliographic System and Services (Кооперативни онлајн библиографски систем и сервиси)

http://www.cobiss.net/platforma_cobiss-SR.htm

CrossRef

<http://www.crossref.org/>

DART Europe

<http://www.dart-europe.eu/basic-search.php>

Dbpedia

<http://dbpedia.org/About>

Deutsche Biographie

<http://www.deutsche-biographie.de/>

DLF– Digital Library Federation

<http://www.diglib.org/>

DPC –Digital Preservation Coallition

<http://www.dpconline.org/>

doiSerbia PhD

<http://www.doiserbia.nb.rs/phd/Default.aspx>

EDM– Europeana Data Model

<http://pro.europeana.eu/edm-documentation>

ESE – Europeana Semantic Elements

<http://pro.europeana.eu/ese-documentation/>

Europeana Portal

http://www.europeana.eu/portal/usingeuropeana_results.html

EuroVoc – Вишејезични речник Европске уније

<http://eurovoc.europa.eu/drupal/?q=sr>

Европске колекције 1914-1918 (Europeana Collections 1914-1918)

<http://www.europeana-collections-1914-1918.eu/>

ExLibris Group. Aleph Integrated Library System

<http://www.exlibrisgroup.com/category/Aleph>

FRBR– *Functional Requirements for Bibliographic Records*

<http://www.ifla.org/publications/functional-requirements-for-bibliographic-records>

Fraunhofer Institute

<http://www.iis.fraunhofer.de/en.html>

Fraunhofer Institute. The Story of MP3: How MP3 was invented

<http://www.mp3-history.com/>

Gemeinsame Normdatei (The Integrated Authority File), доступно на:

http://www.dnb.de/EN/Standardisierung/GND/gnd_node.html

GeoNames

<http://www.geonames.org/>

GIF – Graphic Interchange Format
<http://www.w3.org/Graphics/GIF/spec-gif87.txt>

IBM
<http://www.ibm.com/us-en/>

IFLA – International Federation of Library Association and Institution
<http://www.ifla.org/>

Инфотека
<http://infoteka.bg.ac.rs>

Институт за стандардизацију Србије
<http://www.iss.rs/rs>

International Information Centre for Terminology – Infoterm
<http://www.infoterm.info/>

ISBN – International Standard Book Number
<http://www.isbn.org>

ISMN – International Standard Music Number
<http://www.ismn-international.org/>

ISNI– International Standard Name Identifier
<http://www.isni.org/>

ISSN – International Serial Data Number
<http://www.issn.org/>

JPEG – Joint Photographic Experts Group
<http://www.jpeg.org/about.html>

Корпус савременог српског језика
<http://www.korpus.matf.bg.ac.rs>

LADL – The RELEX Network, Université Paris-Est Marne-la-Vallée
<http://ladl.univ-mlv.fr/Relex/Relex.html>

LCSH– Library of Congress Subject Headings
<http://id.loc.gov/authorities/subjects.html>

Leximir
<http://korpus.matf.bg.ac.rs/soft/LeXimir.html>

Library of Congress
<https://www.loc.gov/>

Linked (Open) Data
<http://linkeddata.org/home>

LISA – Localization Industry Standards Association
<http://web.archive.org/web/20110101184308/http://www.lisa.org/>

LOM– Learning Object Metadata

<http://standards.ieee.org/findstds/standard/1484.12.1-2002.html>

LRMI – Learning resources Metadata Initiative

<http://dublincore.org/dcx/lrmi-terms/>

Marc Code Lists

<http://www.loc.gov/marc/>

MARC Standards

<http://www.loc.gov/marc/>

MarkLogic

<http://www.marklogic.com/>

MemoQ

<https://www.memoq.com/>

METS – Metadata Encoding and Transmission Standard

<http://www.loc.gov/standards/mets/>

МФЦ Микрокомерц

<http://www.earhiva.com/>

Microsoft

<https://www.microsoft.com/sr-latn-rs/>

Microsoft Corporation. WAVE and AVI Codec Registries

<http://tools.ietf.org/html/rfc2361>

Microsoft Corporation. Windows Media Center

<http://windows.microsoft.com/en-US/windows/products/windows-media-center>

MODS – Metadata Object Description Scheme

<http://www.loc.gov/standards/mods/>

MULTEXT-East: Multilingual Text Tools and Corpora for Central and Eastern European Languages

<http://nl.ijs.si/ME/>

Multilingual Thesaurus of the European Union

<http://eurovoc.europa.eu/drupal/>

Народна библиотека Србије. Портал Велики рат

<http://velikirat.nb.rs/onama>

Network Development and MARC Standards Office

<http://www.loc.gov/marc/ndmso.html>

NIBIS - Нишки библиотечко-информациони систем

<http://icit.masfak.ni.ac.rs/sitegenius/topic.php?id=157>

OCLC – Online Computer Library Center

<https://www.oclc.org/home.en.html>

Open Archive Initiative Object Reuse and Exchange

<http://www.openarchives.org/ore/1.0/toc.html>

Open Library

<https://openlibrary.org/>

Österreichische Systematik der Wissenschaftszweige

www.statistik.at/kdb/pages/info_10515.html

OWL – Web Ontology Language

<http://www.w3.org/2001/sw/wiki/OWL>

Oxford English Dictionary

<http://www.oed.com/>

PDF Association. PDF/A facts

<http://www.pdfa.org/2013/02/pdfa-facts/>

PDF Reference (6th ed.)

http://www.adobe.com/content/dam/Adobe/en/devnet/acrobat/pdfs/pdf_reference_1-7.pdf

PHAIDRA – Дигитални репозиторијум Универзитета у Београду

<https://phaidrabg.bg.ac.rs/>

PNG – Portable Network Graphics, Specification (2nd ed.)

<http://www.w3.org/TR/PNG/>

PPT – Microsoft Power Point

<https://products.office.com/en-us/powerpoint>

PREMIS – Preservation Metadata: Implementation Strategies, доступно

на: <http://dx.doi.org/10.7207/twr13-03>

QuickTimeMovie (MOV)

<http://www.apple.com/quicktime/>

RAMEAU– Répertoire d'autorité-matière encyclopédique et alphabétique unifié

<http://rameau.bnf.fr/>

RDF – Resource Description Framework

<http://www.w3.org/RDF/>

RDF Schema

<http://www.w3.org/TR/rdf-schema/>

Semiotischen Thesaurus der Bibliothekwissenschaft

<http://www.ib.hu-berlin.de/~wumsta/infopub/textbook/definitions/thesauindex.html>

SDL Trados

<http://www.translationzone.com/products/sdl-trados-studio/>

Simple Knowledge Organization System

<http://www.w3.org/2009/08/skos-reference/skos.html#>

SQL (Structured Query Language) server

<http://www.microsoft.com/en-us/server-cloud/products/sql-server/>

TBX – Introduction to TermBase eXchange

<http://www.tbxinfo.net/>

TEL – The European Library

<http://www.theeuropeanlibrary.org/tel4/>

TELRI – Trans European Language Resources Infrastructure

<http://telri.nytud.hu/>

Темпус пројекат „Нови библиотечки сервиси на универзитетима Западног Балкана“

<http://www.westbulnet.com/AmplioCMS2/public/EnterPublic.cms2#menuId=1311>

TERMI

<http://termi.rgf.bg.ac.rs/>

Terminological Markup Framework – TMF

<http://www.loria.fr/projets/TMF/>

Tesaurus de Biblioteconomía y Documentación

http://thes.cindoc.csic.es/index_BIBLIO_esp.php

Thomson Reuters

<http://thomsonreuters.com/en.html>

TMX – Translation Memory eXchange– Преводачка меморија за размену

<http://www.gala-global.org/oscarStandards/tmx/tmx14b.html>

UDC – Universal Decimal Classification

<http://www.udcc.org/index.php/site/page?view=about>

UNESCO Thesaurus

<http://databases.unesco.org/thesaurus/>

VIAF - Virtual International Authority File

www.viaf.org

Viet, Jean. Thesaurus Mass Communication. (Paris: UNESCO, 1975)

<http://unesdoc.unesco.org/images/0001/000150/015031MB.pdf>

Виртуелна библиотека Србије

<http://www.vbs.rs/cobiss/>

VebRanka

<http://hlt.rgf.bg.ac.rs/vebranka/>

VebRan

<http://hlt.rgf.bg.ac.rs/vebran/service.asmx>

W3C – World Wide Web Consortium. XML Structure

<http://www.w3pdf.com/W3cSpec/XML/2/REC-xml11-20060816.pdf>

W3C Data Catalog Vocabulary

<http://www.w3.org/TR/vocab-dcat/>

W3C Semantic Web Education and Outreach (SWEO) Interest Group

<http://www.w3.org/blog/SWEO/>

WorldCat

<http://www.oclc.org/worldcat/catalog.en.html>

WorldCat Identities. Univerzitetska biblioteka “Svetozar Marković”

<http://www.worldcat.org/identities/lccn-n80-14562/>

XML – Extensible Markup Language

<http://www.w3.org/XML/>

XML Schema

<http://www.w3.org/standards/xml/schema>

9. Додаци

9.1 Списак слика

Слика 1 – Пример записа за монографску публикацију у формату COMARC/B узајамног каталога COBIB.SR.....	61
Слика 2 – Пример записа за монографску публикацију у формату за експорт података Dublin Core XML	61
Слика 3 – Општи библиографски метаподаци у формату за унос репозиторијума PHAIDRA	65
Слика 4 – Подаци о аутору и ментору докторске дисертације у формату за унос података у репозиторијуму PHAIDRA	66
Слика 5 – Административни метаподаци о врсти лиценце за коришћење дигиталног објекта у формату за унос података у репозиторијуму PHAIDRA.....	67
Слика 6 – Библиографски метаподаци о класификацији у формату за унос података у репозиторијуму PHAIDRA.....	67
Слика 7 – Пример записа за исту докторску дисертацију са припадајућим метаподацима на порталу европских дисертација DART	69
Слика 8 – Пример записа за исту докторску дисертацију у формату за експорт података Dublin Core XML у оквиру репозиторијума PHAIDRA	70
Слика 9 – Дигитална библиотека Филолошког факултета – маска за претрагу преко библиографских метаподатака	73
Слика 10 – Дигитална библиотека Филолошког факултета – маска за претрагу преко структуралних и библиографских метаподатака	74
Слика 11 – Дигитална библиотека Филолошког факултета – пример претраге преко сваке речи у тексту	75
Слика 12 – Претрага на порталу „Европске библиотеке“ – испис по библиотекама у којима се публикација налази	77
Слика 13 – Претрага на порталу „Европске библиотеке“ – графички приказ по земљама у којима су библиотеке које поседују одређену публикацију.....	78
Слика 14 – Пример записа на порталу Европеана у оквиру пројекта „Европске колекције 1914-1918“	79
Слика 15 – Пример записа са слике 14 у дигиталној библиотеци „Велики рат“	80
Слика 16 – Пример записа из збирке оријенталних рукописа на порталу Европеана	81
Слика 17 – Пример записа из збирке оријенталних рукописа у репозиторијуму PHAIDRA	82
Слика 18 – Пример записа из збирке српских новина на порталу „Европска библиотека“	83
Слика 19 – Пример претраге збирке српских новина на претраживачу Универзитетске библиотеке „Светозар Марковић“	84
Слика 20 – Linked Open Data – дијаграм међусобно повезаних скупова података.....	98
Слика 21 – Приказ оптичке сегментације чланака и аутоматског издвајања метаподатака (новине L’echo de Paris, за 24. јун 1937.)	109
Слика 22 – Природно-језичко сучеље за генерисање CQP-упита	113
Слика 23 – Пример постављеног упита у алату Библиша.....	132
Слика 24 – Пример добијених конкорданци у алату Библиша.....	133
Слика 25 – Претрага семантички проширеним упитом у алату Библиша	135
Слика 26 – Пример морфолошког проширења алатом Библиша	135
Слика 27 – Претрага преко метаподатака путем алата Библиша	139
Слика 28 – Резултати претраге преко метаподатака	139
Слика 29 – Преглед паралелног текста чланка преко ТМХ колекције	140
Слика 30 – Задавање претраге термина који се не налази у означеним лексичким ресурсима	141
Слика 31 – Пример претраге термина коришћењем ресурса Библимир	142

Слика 32 – Пример претаге проширивањем упита алатом <i>Библиша</i>	142
Слика 33 – Резултати командног претраживања за тражену кључну реч <i>универзитетска библиотека</i> у каталогу COBIB.SR	145
Слика 34 – Резултати изборног претраживања за тражену кључну реч <i>универзитетска библиотека</i> у каталогу COBIB.SR	145
Слика 35 – Испис комплетних метаподатака у каталогу COBIB.SR	146
Слика 36 – Резултати командног претраживања за тражену кључну реч <i>библиографски опис</i> у каталогу COBIB.SR	148
Слика 37 – Резултати командног претраживања за тражену предметну одредницу <i>архивска грађа</i> у каталогу COBIB.SR	149
Слика 38 – Пример различите садржинске обраде исте публикације у локалним каталозима НБС и УБСМ.....	150
Слика 39 – Приказ резултата командног претраживања у каталогу COBIB.SR преко имена аутора и наслова публикације	151
Слика 40 – Различити облици имена истог аутора у каталогу COBIB.SR	151
Слика 41 – Резултати претраге записа на српском језику на порталу TEL-а на упит <i>универзитетск* библиотек*</i>	152
Слика 42 – Нормативни запис за Универзитетску библиотеку „Светозар Марковић” у бази VIAF	153
Слика 43 – Резултати претраге записа на српском језику на порталу TEL-а на упит <i>библиографск* опис*</i>	155
Слика 44 – Резултати претраге записа на српском језику на порталу TEL-а на упит <i>архивск* грађ*</i> ..	156
Слика 45 – Резултати претраге записа у репозиторијуму PHAIDRA на упит <i>српски језик</i>	158
Слика 46 – Резултати претраге записа у напредном претраживању у репозиторијуму PHAIDRA на упит <i>српски језик</i>	159
Слика 47 – Резултати претраге записа у напредном претраживању у репозиторијуму PHAIDRA на упит <i>корпус савременог српског језика</i>	161
Слика 48 – Маска за претрагу публикација у Дигиталној библиотеци Филолошког факултета	163
Слика 49 – Избор из конкорданци на упит <i>биобиблиографски речник</i> у Дигиталној библиотеци Филолошког факултета.....	164
Слика 50 – Комбиновано претраживање преко дескриптора метаподатака и дескриптора садржаја у Дигиталној библиотеци Филолошког факултета	166
Слика 51 – Део погодака на упит <i>текућа библиографија</i> у Дигиталној библиотеци Филолошког факултета	167
Слика 52 – Делимична листа различитих облика имена Александар Велики у бази ISNI.....	169
Слика 53 – Пример претраге новинске колекције упитом <i>Бранко Радичевић</i> са делом конкорданци	173
Слика 54 – Пример једног резултата претраге новинске колекције на упит <i>Бранко Радичевић</i>	174
Слика 55 – Комплетан текст чланка са препознатим речима из упита	175
Слика 56 – Грешке у OCR-у – непрепознати латинични карактери у ћириличном тексту.....	176
Слика 57 – Оригинални текст са латиничним писмом у ћириличном тексту	176
Слика 58 – Пример претраге новинске колекције где је пронађена само једна реч из упита	177
Слика 59 – Пример погодака у новинској колекцији који се односе на националне библиотеке	178
Слика 60 – Пример непрепознавања различитих морфолошких облика у претрази новинске колекције.....	178
Слика 61 – Пример записа са метаподацима на порталу Велики рат	181
Слика 62 – Веза ка бази GeoNames из записа на порталу Велики рат	181
Слика 63 – Претрага преко метаподатака у дигиталној библиотеци <i>Библиша</i>	183
Слика 64 – Пример претраге кључном речју <i>универзитетска библиотека</i> у дигиталној библиотеци <i>Библиша</i>	183

Слика 65 – Избор из конкорданци добијених на упит <i>универзитетска библиотека</i> у дигиталној библиотеци <i>Библиша</i>	184
Слика 66 – Избор из конкорданци добијених на упит <i>мултимедијални документ</i> у дигиталној библиотеци <i>Библиша</i>	185
Слика 67 – Нивои термилошке одреднице	208
Слика 68 – Маска за унос новог термина у базу Терми	211
Слика 69 – Маска за унос библиографских података о извору у бази Терми	212
Слика 70 – Извоз термилошких података из базе Терми у TBX формату	213
Слика 71 – Архитектура система	220
Слика 72 – Прецизност проналажења 1) вишечлане речи; 2) леме; 3) термина	222

9.2 Списак табела

Табела 1 – Елементи модела EDM	91
Табела 2 – Листа морфолошких категорија и кодова у морфолошком речнику српског језика	116
Табела 3 – Приказ маркера придружених лемама у српском DELAS речнику	117
Табела 4 – Дескриптори речи са објашњењима у „Библиотекарском термилошком речнику“	122
Табела 5 – Кораци у производњи леме за вишечлану реч	217
Табела 6 – Рангирање пронађених облика и лема вишечланих речи	219
Табела 7 – Број лема са више од 3 појављивања у доменском корпусу	221
Табела 8 – Најфреквентније вишечлане леме у доменском корпусу	226
Табела 9 – Фреквенција именованих ентитета према врстама у доменском корпусу	234
Табела 10 – Вишечлани термини препознати у новом тексту	237
Табела 11 – Кандидати за термилошки речник са највећом фреквенцијом у новом корпусу	239
Табела 12 – Фреквенција именованих ентитета према врстама у новом корпусу	239

Прилози

Прилог 1 - Вишечлане речи одабране за допуну термилошког речника из библиотекарства и информатике

Obrazac	graf	broj	lema	FrekGrafLema	EvalTerm
AXN	grf01	sin	aerodromska biblioteka	9	1
N2X	grf03	sin	agregator podataka	4	1
AXN2X	grf05	sin	akademska baza podataka	7	1
AXN	grf01	sin	akademska biblioteka	23	1
N4X	grf04a	sin	alat za obradu	10	1
AXN	grf01	sin	alfabetski katalog	16	1
AXN	grf01	sin	alfabetski princip	4	1
AXN	grf01	sin	algoritamski pristup	5	1
AXN	grf01	sin	analitička bibliografija	33	1
AXN	grf01	sin	analitička obrada	4	1
N2X	grf03	sin	analiza sadržaja	4	1
N2X	grf03	sin	analiza teksta	4	1
N2X	grf03	sin	anotacija korpusa	12	1
AXN	grf01	sin	arhivska građa	40	1
2XN	grf02	sin	audio-datoteka	6	1
2XAXN	grf10	sin	audio-vizuelni materijal	4	1
AXN	grf01	sin	automatska analiza	4	1
AXN	grf01	sin	automatska anotacija	6	1
AXN	grf01	sin	automatska obrada	11	1
N2X	grf03	sin	autor dela	4	1
AXN	grf01	sin	autorsko delo	11	1
AXN	grf01	sin	azbučni red	8	1
AXN	grf01	sin	azbučni redosled	7	1
N2X	grf03	sin	banka stabala	5	1
AXN4X	grf09a	sin	bela knjiga o jezicima	4	1
N2X	grf03	sin	bibliografija knjiga	9	1
N2X	grf03	sin	bibliografija periodike	6	1
AXN	grf01	sin	bibliografska agencija	4	1
AXN	grf01	sin	bibliografska anotacija	4	1
AXN	grf01	sin	bibliografska baza	10	1
AXN	grf01	sin	bibliografska beleška	12	1
AXN	grf01	sin	bibliografska građa	10	1
AXN	grf01	sin	bibliografska informacija	9	1
AXN	grf01	sin	bibliografska jedinica	81	1
AXN	grf01	sin	bibliografska kontrola	37	1
AXN	grf01	sin	bibliografska manšeta	4	1
AXN	grf01	sin	bibliografska obrada	16	1
AXN2X	grf05	sin	bibliografska obrada podataka	4	1
AXN	grf01	sin	bibliografska praksa	7	1
AXN	grf01	sin	bibliografska studija	7	1
AXN	grf01	sin	bibliografska tehnika	5	1
AXN	grf01	sin	bibliografski centar	7	1
AXN	grf01	sin	bibliografski opis	87	1

AXN	grf01	plu	bibliografski podaci	46	1
AXN	grf01	plu	bibliografski pregledi	9	1
AXN	grf01	sin	bibliografski prilog	7	1
AXN	grf01	sin	bibliografski priručnik	4	1
AXN	grf01	sin	bibliografski rad	12	1
AXN	grf01	sin	bibliografski zapis	8	1
AXN	grf01	sin	bibliografsko citiranje	11	1
AXN	grf01	sin	bibliografsko istraživanje	10	1
AXN	grf01	sin	bibliografsko opisivanje	22	1
AXN	grf01	sin	bibliotečka asocijacija	5	1
AXN	grf01	sin	bibliotečka delatnost	107	1
AXN	grf01	sin	bibliotečka građa	95	1
AXN	grf01	sin	bibliotečka jedinica	7	1
AXN	grf01	sin	bibliotečka mreža	13	1
AXN2X	grf05	sin	bibliotečka mreža univerziteta	4	1
AXN	grf01	sin	bibliotečka praksa	6	1
AXN	grf01	sin	bibliotečka profesija	4	1
AXN	grf01	sin	bibliotečka služba	6	1
AXN	grf01	sin	bibliotečka statistika	18	1
AXN	grf01	sin	bibliotečka zajednica	9	1
AXN	grf01	sin	bibliotečka zbirka	5	1
AXN	grf01	sin	bibliotečka zgrada	4	1
AXN	grf01	plu	bibliotečke aktivnosti	4	1
AXN	grf01	plu	bibliotečke kolekcije	5	1
AXN	grf01	sin	bibliotečki centar	9	1
AXN	grf01	sin	bibliotečki fond	33	1
AXN	grf01	plu	bibliotečki izvori	4	1
AXN	grf01	sin	bibliotečki kadar	12	1
AXN	grf01	plu	bibliotečki konzorcijumi	4	1
AXN	grf01	sin	bibliotečki materijal	70	1
AXN	grf01	sin	bibliotečki odbor	8	1
AXN	grf01	plu	bibliotečki poslovi	15	1
AXN	grf01	plu	bibliotečki servisi	11	1
AXN	grf01	sin	bibliotečki sistem	42	1
AXN	grf01	sin	bibliotečko osoblje	20	1
AXN	grf01	sin	bibliotečko poslovanje	41	1
2XAXN	grf10	sin	bibliotečko-bibliografska klasifikacija	4	1
2XAXN	grf10	sin	bibliotečko-informaciona delatnost	13	1
2XAXN	grf10	sin	bibliotečko-informacioni sistem	47	1
NXN	grf06	sin	biblioteka celina	12	1
NXN	grf06	sin	biblioteka članica	7	1
N2X	grf03	sin	biblioteka fakulteta	10	1
N4X	grf04a	sin	biblioteka na aerodromu	4	1
N4X	grf04a	sin	biblioteka na brodu	7	1
N4X	grf04a	sin	biblioteka na univerzitetu	17	1
N4X	grf04a	sin	biblioteka u metrou	9	1
NXN	grf06	sin	biblioteka-celina	23	1
NXN	grf06	sin	biblioteka-članica	5	1

AXN	grf01	plu	biografski podaci	14	1
AXN2X	grf05	sin	brzi protok informacija	4	1
AXN	grf01	sin	centralna biblioteka	39	1
AXN	grf01	sin	centralni katalog	39	1
AXN	grf01	sin	centralni registar	17	1
AXN4X	grf09b	sin	centralni registar narodne biblioteke	13	1
AXN	grf01	sin	ciljna grupa	5	1
AXN	grf01	sin	ćirilčno pismo	6	1
AXN	grf01	sin	ćirilska knjiga	4	1
AXN	grf01	sin	čitalačka publika	32	1
AXN	grf01	sin	citatna baza	4	1
N2X	grf03	sin	čuvanje građe	4	1
N2X	grf03	sin	čuvanje podataka	8	1
AXN	grf01	sin	dablinsko jezgro	10	1
N2X	grf03	sin	datum štampanja	4	1
N2X	grf03	sin	delatnost biblioteke	9	1
N2X	grf03	sin	digitalizacija knjiga	4	1
AXN	grf01	sin	diskusiona grupa	5	1
N2X	grf03	sin	distribucija informacija	5	1
AXN	grf01	sin	dnevna novina	4	1
AXN	grf01	sin	dnevna štampa	4	1
AXN	grf01	plu	dopunski elementi	6	1
NXN	grf06	sin	dostupnost informacija	5	1
N2X	grf03	sin	dostupnost publikacija	11	1
N2X	grf03	sin	dostupnost znanja	4	1
N2X	grf03	sin	društvo znanja	4	1
AXN	grf01	sin	državna biblioteka	12	1
AXN	grf01	sin	državna štamparija	7	1
AXN2X	grf05	sin	dugoročno čuvanje podataka	5	1
N2X	grf03	sin	ekstrakcija informacija	23	1
AXN	grf01	sin	elektronska baza	14	1
AXN2X	grf05	sin	elektronska baza podataka	11	1
AXN	grf01	sin	elektronska biblioteka	5	1
AXN	grf01	sin	elektronska forma	5	1
AXN	grf01	sin	elektronska građa	4	1
AXN	grf01	sin	elektronska knjiga	5	1
AXN	grf01	sin	elektronska periodika	4	1
AXN	grf01	sin	elektronska pošta	12	1
AXN	grf01	plu	elektronske usluge	4	1
AXAXN	grf07	plu	elektronski informacioni izvori	4	1
AXN	grf01	plu	elektronski mediji	5	1
AXN	grf01	sin	elektronski rečnik	14	1
AXN	grf01	sin	elektronsko izdanje	10	1
AXN	grf01	sin	elektronsko izdavanje	4	1
AXN	grf01	sin	elektronsko učenje	43	1
N4X	grf04b	sin	etiketiranje vrstom reči	8	1
AXN	grf01	sin	etimološko poreklo	6	1
N4X	grf04b	sin	evaluacija bibliotečkih fondova	4	1

N4X	grf04a	sin	evaluacija u bibliotekama	4	1
AXN	grf01	sin	evropska biblioteka	10	1
AXN	grf01	sin	fakultetska biblioteka	34	1
N2X	grf03	sin	fond biblioteke	27	1
N2X	grf03	sin	format resursa	4	1
N2X	grf03	sin	frekvencija korišćenja	4	1
N2X	grf03	sin	funkcija biblioteke	8	1
AXN	grf01	sin	geografska karta	5	1
AXN	grf01	sin	geografska lokacija	7	1
AXN	grf01	sin	geološka terminologija	10	1
AXAXN	grf07	sin	geološki informacioni sistem	9	1
AXN	grf01	sin	geološki rečnik	8	1
N2X	grf03	sin	godina izdavanja	16	1
N2X	grf03	sin	godina izlaženja	9	1
N2X	grf03	sin	godina štampanja	5	1
AXN2X	grf05	plu	govorne baze podataka	4	1
AXN	grf01	plu	govorne tehnologije	4	1
AXN	grf01	sin	govorni korpus	11	1
N4X	grf04a	sin	građa za bibliografiju	10	1
AXN	grf01	sin	građanska ćirilica	4	1
AXN	grf01	sin	gradska biblioteka	16	1
AXN	grf01	sin	grafička mera	4	1
AXN	grf01	sin	gramatička informacija	4	1
N6X	grf08c	sin	grupa za jezičke tehnologije	15	1
AXN	grf01	sin	grupna raspodela	51	1
AXN	grf01	sin	hijerarhijska struktura	6	1
AXN	grf01	sin	hronološki kriterijum	6	1
AXN	grf01	sin	hronološki princip	11	1
AXN	grf01	sin	hronološki raspored	4	1
AXN	grf01	sin	hronološki redosled	8	1
AXN	grf01	sin	hronološki sled	5	1
N2X	grf03	sin	ime autora	56	1
N2X	grf03	sin	ime pisca	7	1
N2X	grf03	sin	ime prevodioca	12	1
N2X	grf03	sin	ime saradnika	4	1
N4X	grf04a	sin	ime u nominativu	5	1
AXN	grf01	sin	imenička fraza	6	1
AXN	grf01	sin	imenski prostor	6	1
N4X	grf04a	plu	informacije o izvorima	4	1
AXN	grf01	sin	informaciona pismenost	11	1
AXN	grf01	sin	informacioni centar	10	1
AXN2X	grf05	sin	informacioni sistem univerziteta	9	1
AXN	grf01	sin	informaciono društvo	16	1
2XAXN	grf10	sin	informaciono-bibliotečki centar	4	1
AXN	grf01	sin	informatičko društvo	7	1
N2X	grf03	sin	informisanje korisnika	6	1
N6X	grf08c	sin	inicijativa za kodiranje teksta	8	1
AXN	grf01	sin	inostrana literatura	6	1

N2X	grf03	sin	institucija kulture	7	1
AXN2X	grf05	sin	intelektualna svojina autora	4	1
AXN	grf01	sin	inventarska jedinica	4	1
AXN	grf01	sin	inventarska knjiga	18	1
N2X	grf03	sin	isporuka dokumenata	7	1
N2X	grf03	sin	istorija bibliografije	6	1
AXN	grf01	sin	istorijska biblioteka	4	1
AXN	grf01	sin	istorijska perspektiva	8	1
AXN	grf01	sin	istorijski arhiv	6	1
AXN	grf01	sin	istraživački proces	18	1
AXN	grf01	sin	istraživački rad	42	1
AXN	grf01	sin	izdavačka celina	8	1
AXN	grf01	sin	izdavačka delatnost	34	1
AXN	grf01	sin	izdavačka kuća	8	1
AXN	grf01	sin	izdavačka produkcija	50	1
NXN	grf06	sin	izdavanje publikacija	4	1
AXN	grf01	sin	jezička tehnologija	43	1
AXN	grf01	sin	jezički model	6	1
AXN	grf01	sin	jezički resurs	5	1
N4X	grf04a	sin	jezik za obeležavanje	8	1
AXN	grf01	sin	jugoslovenska bibliografija	10	1
AXAXN	grf07	sin	jugoslovenski bibliografski institut	21	1
AXN	grf01	sin	kadrovska politika	6	1
AXN	grf01	sin	kalendarska književnost	6	1
AXN	grf01	sin	kancelarijski paket	5	1
AXN	grf01	sin	kartografska građa	8	1
N2X	grf03	sin	katalog knjiga	5	1
N4X	grf04a	sin	katalogizacija na izvoru	8	1
N4X	grf04a	sin	katalogizacija u publikaciji	9	1
AXN	grf01	sin	kataloška jedinica	17	1
AXN	grf01	sin	kataloška obrada	6	1
AXN	grf01	sin	kataloška praksa	9	1
AXN	grf01	plu	kataloška pravila	11	1
AXN	grf01	sin	kataloški listić	6	1
AXN	grf01	sin	kataloški opis	7	1
AXN	grf01	sin	kataloški zapis	4	1
AXN	grf01	sin	katedarska biblioteka	4	1
N2X	grf03	sin	klasa etiketa	6	1
N4X	grf04b	sin	klasifikacija bibliotečkog materijala	5	1
N2X	grf03	sin	klasifikacija građe	4	1
AXN	grf01	plu	klasifikacione sheme	7	1
AXN	grf01	sin	ključna reč	48	1
N2X	grf03	sin	knjiga inventara	4	1
AXN	grf01	sin	knjižna produkcija	6	1
AXN	grf01	sin	knjižni fond	13	1
N2X	grf03	sin	kodiranje teksta	9	1
AXN	grf01	sin	kodni raspored	4	1
AXN2X	grf05	plu	komercijalne baze podataka	8	1

N2X	grf03	sin	kontrola kvaliteta	7	1
N2X	grf03	sin	konverzija teksta	8	1
N6X	grf08a	sin	konverzija teksta u govor	8	1
N4X	grf04b	sin	konzervacija bibliotečke građe	4	1
N4X	grf04a	plu	koraci u programiranju	5	1
N2X	grf03	sin	korišćenje biblioteke	8	1
N4X	grf04b	sin	korišćenje elektronskih izvora	9	1
N2X	grf03	sin	korišćenje informacija	8	1
N2X	grf03	sin	korišćenje podataka	9	1
N2X	grf03	sin	korišćenje resursa	10	1
AXN	grf01	sin	korisnička grupa	7	1
AXN	grf01	plu	korisničke potrebe	10	1
AXN	grf01	sin	korisničko ime	7	1
N2X	grf03	sin	korisnik biblioteke	18	1
N2X	grf03	sin	korpus tekstova	4	1
N4X	grf04a	sin	korpus za obučavanje	14	1
N4X	grf04a	sin	korpus za obuku	6	1
N4X	grf04a	sin	korpus za testiranje	4	1
AXN	grf01	sin	korpusna lingvistika	9	1
AXN	grf01	sin	korpusna reč	9	1
AXN	grf01	sin	kraljevska biblioteka	8	1
AXN	grf01	sin	kritička bibliografija	4	1
AXN	grf01	sin	kritički aparat	9	1
AXN	grf01	sin	kritičko izdanje	9	1
N2X	grf03	sin	kvalitet usluga	10	1
AXN	grf01	sin	latinično pismo	8	1
AXN	grf01	sin	leksička baza	17	1
AXN2X	grf05	sin	leksička baza podataka	14	1
AXN	grf01	sin	leksički obrazac	9	1
AXN	grf01	sin	leksički resurs	6	1
AXN	grf01	plu	leksički resursi	10	1
AXN	grf01	sin	leksičko prepoznavanje	10	1
N4X	grf04a	sin	licenca za rad	8	1
AXN	grf01	plu	ljudski resursi	6	1
AXN	grf01	sin	logička struktura	13	1
N4X	grf04b	sin	lokalizacija slobodnog softvera	4	1
AXN	grf01	sin	magacinski prostor	4	1
AXN	grf01	sin	magistarski rad	13	1
N4X	grf04a	sin	mašina za pretraživanje	4	1
AXN	grf01	sin	mašinska obrada	4	1
AXN	grf01	sin	mašinsko prevođenje	33	1
AXN	grf01	sin	mašinsko učenje	21	1
N4X	grf04a	sin	materijal za učenje	14	1
AXN	grf01	sin	matična biblioteka	16	1
AXN	grf01	sin	matična funkcija	6	1
AXN	grf01	sin	matična institucija	6	1
AXN	grf01	sin	međubibliotečka pozajmica	61	1
AXN	grf01	sin	međubibliotečka razmena	5	1

AXN	grf01	sin	međubibliotečka saradnja	6	1
AXN	grf01	sin	međunarodna bibliografija	17	1
AXN	grf01	sin	međunarodna konferencija	13	1
AXN	grf01	sin	međunarodna organizacija	14	1
AXN	grf01	sin	međunarodna razmena	6	1
AXN	grf01	sin	međunarodna saradnja	11	1
AXN	grf01	sin	međunarodni karakter	8	1
AXN	grf01	sin	međunarodni standard	20	1
AXN	grf01	sin	metodičko uputstvo	4	1
N2X	grf03	sin	misija biblioteke	4	1
N2X	grf03	sin	mobilnost stručnjaka	4	1
N2X	grf03	sin	model podataka	13	1
N4X	grf04a	sin	modul za upravljanje	4	1
AXN	grf01	sin	monografska publikacija	39	1
AXN	grf01	sin	morfološka analiza	12	1
AXN	grf01	sin	morfološka obrada	4	1
AXAXN	grf07	sin	morfološki elektronski rečnik	4	1
AXN	grf01	sin	morfološki rečnik	14	1
AXN	grf01	sin	morfološko proširenje	9	1
N2X	grf03	sin	mreža biblioteka	11	1
N2X	grf03	sin	mreža reči	26	1
AXN	grf01	sin	mrežni nivo	6	1
AXN	grf01	sin	muzejska građa	4	1
N4X	grf04b	sin	nabavka bibliotečke građe	17	1
N4X	grf04b	sin	nabavka bibliotečkog materijala	5	1
N2X	grf03	sin	nabavka literature	6	1
AXN	grf01	sin	nabavna politika	27	1
N2X	grf03	sin	nacrt standard	9	1
AXN	grf01	sin	nadzorni odbor	12	1
AXN	grf01	sin	narodna biblioteka	192	1
N2X	grf03	sin	naslov časopisa	9	1
N2X	grf03	sin	naslov dela	10	1
N2X	grf03	sin	naslov publikacije	4	1
AXN	grf01	sin	naslovni list	4	1
AXN	grf01	sin	naučna baština	6	1
AXN	grf01	sin	naučna biblioteka	12	1
AXN	grf01	sin	naučna disciplina	32	1
AXN	grf01	sin	naučna grana	6	1
AXN	grf01	sin	naučna javnost	7	1
AXN	grf01	sin	naučna literature	10	1
AXN	grf01	sin	naučna metoda	9	1
AXN	grf01	sin	naučna oblast	28	1
AXN	grf01	sin	naučna zajednica	15	1
AXN	grf01	sin	naučni document	5	1
AXN	grf01	sin	naučni metod	6	1
AXN	grf01	sin	naučni rad	62	1
AXN	grf01	sin	naučni radnik	7	1
AXN	grf01	sin	naučno istraživanje	21	1

AXN	grf01	sin	naučno izdanje	6	1
AXN	grf01	sin	naučno izdavaštvo	4	1
AXN	grf01	plu	naučnoistraživačke organizacije	6	1
2XN	grf02	sin	n-gram	4	1
N2X	grf03	sin	obeležavanje teksta	5	1
N2X	grf03	sin	objavljivanje publikacije	7	1
2XAXN	grf10	sin	objektno-orijentisano programiranje	4	1
AXN	grf01	sin	oblasna biblioteka	4	1
N4X	grf04b	sin	oblast ljudskog znanja	8	1
N2X	grf03	sin	oblast nauke	6	1
N2X	grf03	sin	oblast znanja	5	1
N2X	grf03	sin	oblik reči	26	1
N4X	grf04b	sin	obrada bibliotečkog materijala	4	1
N2X	grf03	sin	obrada jezika	8	1
N4X	grf04b	sin	obrada prirodnih jezika	22	1
N2X	grf03	sin	obrazovanje bibliotekara	6	1
AXN	grf01	sin	obrazovna institucija	6	1
AXN	grf01	sin	obrazovni proces	17	1
N2X	grf03	sin	obučavanje korisnika	6	1
N4X	grf04b	sin	očuvanje kulturnog nasleđa	6	1
N4X	grf04a	sin	odgovaranje na pitanja	10	1
AXN	grf01	sin	ontološka kategorija	10	1
NXN	grf06	sin	opis knjiga	5	1
NXN	grf06	sin	opis publikacija	4	1
N2X	grf03	sin	opis rukopisa	8	1
N4X	grf04b	sin	opis višetomnih publikacija	4	1
N2X	grf03	sin	opisivanje građe	5	1
AXN2X	grf05	sin	optičko prepoznavanje karaktera	8	1
N2X	grf03	sin	organizacija biblioteke	6	1
AXN	grf01	sin	organizaciona jedinica	10	1
AXN	grf01	sin	organizaciona struktura	11	1
N2X	grf03	sin	osoblje biblioteke	5	1
N2X	grf03	sin	periodičnost izlaženja	10	1
AXN	grf01	sin	piratski softver	4	1
N2X	grf03	sin	pismo originala	11	1
AXN2X	grf05	sin	početna godina izlaženja	5	1
N4X	grf04a	plu	podaci o autoru	7	1
N4X	grf04a	plu	podaci o citiranosti	7	1
N4X	grf04a	plu	podaci o mestu	6	1
N4X	grf04a	plu	podaci o naslovu	8	1
N4X	grf04a	plu	podaci o odgovornosti	13	1
N4X	grf04a	sin	podatak o izdanju	16	1
N6X	grf08c	sin	podatak o periodičnosti izlaženja	4	1
N4X	grf04a	sin	podatak o pismu	4	1
N4X	grf04a	sin	podatak o tiražu	5	1
N2X	grf03	sin	područje napomena	4	1
NXN	grf06	sin	popis publikacija	5	1
N2X	grf03	sin	poslovanje biblioteke	12	1

N4X	grf04a	plu	poslovi u biblioteci	6	1
AXN	grf01	sin	pozajmno mesto	6	1
AXN	grf01	sin	pravni status	7	1
N2X	grf03	sin	prečišćavanje teksta	4	1
N2X	grf03	sin	predmet bibliografije	19	1
N2X	grf03	sin	predmet istraživanja	10	1
N2X	grf03	sin	predmet publikacije	6	1
AXN	grf01	sin	predmetna odrednica	7	1
AXN	grf01	sin	predmetna specijalizacija	5	1
AXN4X	grf09b	plu	predmetne odrednice kongresne	4	1
AXN	grf01	sin	predmetni katalog	9	1
AXN	grf01	sin	predmetni specijalista	10	1
N2X	grf03	sin	prenos informacija	5	1
N2X	grf03	sin	prepoznavanje govora	6	1
N4X	grf04b	sin	prepoznavanje imenovanih entiteta	15	1
N2X	grf03	sin	prepoznavanje karaktera	10	1
N2X	grf03	sin	prezentacija dokumenta	6	1
N2X	grf03	sin	prezime autora	10	1
N2X	grf03	sin	prikaz sadržaja	25	1
N2X	grf03	sin	prikupljanje podataka	12	1
N2X	grf03	sin	primerak publikacije	6	1
N2X	grf03	sin	proces nabavke	6	1
N2X	grf03	sin	proces učenja	10	1
N4X	grf04a	sin	program za etiketiranje	28	1
AXN	grf01	sin	programski savet	8	1
N2X	grf03	sin	proizvođač softvera	5	1
AXN	grf01	sin	projektni tim	4	1
N2X	grf03	sin	pronalaženje informacija	52	1
N2X	grf03	sin	pronalaženje podataka	4	1
AXN	grf01	sin	prosvetna delatnost	7	1
N2X	grf03	sin	protok podataka	4	1
N2X	grf03	sin	provajder podataka	21	1
AXN	grf01	sin	računarska laboratorija	5	1
AXN	grf01	sin	računarska lingvistika	27	1
AXN	grf01	sin	računarska mreža	7	1
AXN	grf01	sin	računarska oprema	13	1
AXN	grf01	sin	računarska tehnologija	17	1
N4X	grf04a	sin	računarstvo u oblaku	9	1
AXN	grf01	sin	računski centar	5	1
AXN	grf01	sin	radna grupa	28	1
AXN	grf01	sin	radna organizacija	6	1
AXN	grf01	plu	radne obaveze	5	1
AXN	grf01	sin	radni paket	6	1
AXN	grf01	sin	radni proces	6	1
AXN	grf01	sin	radni prostor	4	1
AXN	grf01	sin	radno mesto	12	1
AXN	grf01	sin	radno okruženje	4	1
N2X	grf03	sin	raspored građe	4	1

N2X	grf03	sin	razmena podataka	7	1
N2X	grf03	sin	razmena poruka	4	1
N2X	grf03	sin	razmena resursa	5	1
N2X	grf03	sin	razvoj sistema	10	1
N2X	grf03	sin	razvoj softvera	7	1
N2X	grf03	sin	razvoj štamparstva	4	1
AXN	grf01	sin	razvojno okruženje	8	1
AXN	grf01	sin	rečnička odrednica	12	1
N2X	grf03	sin	red reči	6	1
AXN2X	grf05	sin	redni broj izdanja	4	1
AXN	grf01	sin	referensna služba	6	1
AXN	grf01	sin	referensna usluga	4	1
NXN	grf06	sin	registar biblioteka	6	1
AXN	grf01	sin	relacioni pridev	6	1
N2X	grf03	sin	rezultat istraživanja	9	1
N2X	grf03	sin	rezultat pretrage	6	1
AXN	grf01	sin	rukopisna građa	16	1
AXN	grf01	sin	rukopisna knjiga	15	1
AXN	grf01	sin	rukopisni materijal	4	1
N2X	grf03	sin	sadržaj dokumenta	4	1
N2X	grf03	sin	sadržaj publikacije	4	1
N4X	grf04a	sin	saradnik na bibliografiji	5	1
N2X	grf03	sin	saradnja biblioteka	5	1
AXN	grf01	sin	sastavni deo	34	1
N2X	grf03	sin	selekcija građe	4	1
AXN2X	grf05	sin	selektivna diseminacija informacija	7	1
AXN	grf01	sin	semantička mreža	21	1
AXN2X	grf05	sin	semantička mreža reči	8	1
AXN	grf01	sin	semantička povezanost	4	1
AXN	grf01	sin	semantička relacija	5	1
AXN	grf01	sin	semantička skala	5	1
AXN	grf01	sin	semantička veza	5	1
AXN	grf01	sin	semantički opis	4	1
AXN	grf01	sin	semantički veb	35	1
AXN	grf01	sin	seminarska biblioteka	22	1
AXN	grf01	sin	seminarski rad	13	1
AXN	grf01	sin	serijska publikacija	55	1
N4X	grf04a	sin	servis u oblaku	5	1
AXN	grf01	sin	signalna informacija	4	1
AXN	grf01	sin	sintaksička struktura	8	1
N2X	grf03	sin	sistem kataloga	4	1
N2X	grf03	sin	sistem kodiranja	4	1
N2X	grf03	sin	sistem kvaliteta	21	1
N4X	grf04b	sin	sistem morfoloških rečnika	6	1
N2X	grf03	sin	sistem upravljanja	8	1
N4X	grf04b	sin	sistem uzajamne katalogizacije	21	1
N4X	grf04a	sin	sistem za ekstrakciju	5	1
N4X	grf04a	sin	sistem za obeležavanje	4	1

N4X	grf04a	sin	sistem za pretraživanje	4	1
N6X	grf08c	sin	sistem za pronalaženje informacija	13	1
N2X	grf03	sin	sistematizacija građe	14	1
N4X	grf04b	sin	sistematizacija ljudskog znanja	4	1
AXN	grf01	sin	školska biblioteka	8	1
AXN	grf01	sin	slovensko pismo	10	1
N2X	grf03	sin	snabdevanje dokumentima	9	1
N4X	grf04b	sin	snabdevanje naučnim dokumentima	7	1
N4X	grf04b	sin	softver otvorenog koda	15	1
AXN	grf01	sin	softverski sistem	5	1
AXN	grf01	sin	softversko rešenje	8	1
N2X	grf03	sin	spomenik kulture	10	1
N6X	grf08c	sin	sporazum o razmeni podataka	8	1
AXN	grf01	sin	srpska bibliografija	100	1
AXN	grf01	sin	srpska knjiga	37	1
AXN	grf01	sin	srpska periodika	11	1
AXN	grf01	sin	srpska štampa	6	1
AXN	grf01	plu	srpske biblioteke	4	1
AXN	grf01	plu	srpske novine	12	1
AXN	grf01	plu	srpske štamparije	4	1
AXN	grf01	plu	srpski autori	7	1
AXN	grf01	plu	srpski bibliografi	5	1
AXN	grf01	plu	srpski časopisi	4	1
AXN	grf01	plu	srpski kalendari	6	1
AXN	grf01	sin	srpsko štamparstvo	6	1
AXN	grf01	sin	štamparska delatnost	8	1
AXN	grf01	sin	štamparsko slovo	4	1
AXN	grf01	sin	statistički metod	5	1
AXN	grf01	plu	statistički podaci	43	1
AXAXN	grf07	sin	statističko mašinsko prevođenje	6	1
N4X	grf04b	sin	stil bibliografskog citiranja	12	1
N4X	grf04a	sin	stranica za razgovor	8	1
AXN	grf01	sin	strateški plan	11	1
N2X	grf03	sin	struktura podataka	7	1
N2X	grf03	sin	studija slučaja	5	1
N4X	grf04a	sin	sumeđa na vebu	4	1
AXN	grf01	sin	svetska mreža	16	1
AXN	grf01	sin	sveučilišna knjižnica	12	1
AXN	grf01	sin	tehnička dokumentacija	4	1
AXN	grf01	sin	tehnička podrška	8	1
AXN	grf01	plu	tehničke informacije	8	1
AXN	grf01	plu	tehničke mogućnosti	4	1
AXN	grf01	plu	tehničke osobine	4	1
AXN	grf01	sin	tehničko osoblje	6	1
N2X	grf03	sin	tehnologija oblaka	37	1
AXN	grf01	sin	tehnološka inovacija	6	1
AXN	grf01	sin	tehnološka podrška	4	1
AXN	grf01	sin	tehnološki napredak	10	1

AXN	grf01	sin	tehnološki razvoj	11	1
AXN	grf01	sin	tematska bibliografija	4	1
N2X	grf03	sin	teoretičar bibliotekarstva	8	1
N2X	grf03	sin	teorija bibliografije	15	1
N2X	grf03	sin	teorija bibliotekarstva	6	1
AXN	grf01	sin	terminološki rečnik	6	1
AXN	grf01	plu	terminološki resursi	9	1
AXN	grf01	sin	timski rad	8	1
N2X	grf03	sin	tip biblioteke	8	1
N2X	grf03	sin	tipologija biblioteka	4	1
AXN	grf01	sin	topografski katalog	4	1
N4X	grf04a	sin	učenje na daljinu	32	1
AXN	grf01	sin	udžbenička literatura	11	1
N2X	grf03	sin	uklanjanje sufiksa	14	1
N2X	grf03	sin	uloga bibliotekara	7	1
N2X	grf03	sin	uloga biblioteke	6	1
AXN	grf01	sin	umetničko delo	4	1
N2X	grf03	sin	unifikacija opisivanja	4	1
AXN2X	grf05	sin	univerzalna dostupnost publikacija	9	1
AXN2X	grf05	sin	univerzalni protok podataka	4	1
AXN	grf01	sin	univerzitetska biblioteka	373	1
AXN	grf01	sin	univerzitetski bibliotekar	4	1
N2X	grf03	sin	unos podataka	8	1
AXN	grf01	sin	upravljačka struktura	4	1
N2X	grf03	sin	upravljanje informacijama	4	1
N2X	grf03	sin	upravljanje sadržajem	32	1
N4X	grf04b	sin	upravljanje višejezičnim sadržajima	6	1
N2X	grf03	sin	upravljanje znanjem	6	1
N2X	grf03	sin	upravnik biblioteke	7	1
N4X	grf04a	sin	uputstvo za korišćenje	7	1
N2X	grf03	sin	uređivanje teksta	4	1
N2X	grf03	sin	uredništvo časopisa	5	1
N2X	grf03	sin	usluga biblioteke	6	1
NXN	grf06	sin	vab aplikacija	13	1
NXN	grf06	sin	vab sajt	5	1
NXN	grf06	sin	vab servis	4	1
NXN	grf06	sin	vab stanica	6	1
AXN	grf01	sin	vektorski model	4	1
N2X	grf03	sin	veličina slova	4	1
AXN	grf01	sin	veštačka inteligencija	4	1
AXN	grf01	sin	visokoškolska biblioteka	95	1
AXN	grf01	sin	visokoškolska ustanova	9	1
AXN	grf01	sin	visokoškolsko obrazovanje	7	1
AXN2X	grf05	sin	vizuelna prezentacija dokumenta	6	1
N2X	grf03	sin	vlasnik biblioteke	5	1
N2X	grf03	sin	vrednost atributa	6	1
AXN	grf01	sin	vremenski period	19	1
AXN	grf01	sin	vremenski raspon	7	1

N2X	grf03	sin	vrsta bibliografije	4	1
N4X	grf04b	sin	vrsta bibliotečke građe	10	1
N2X	grf03	sin	vrsta dokumenata	6	1
N2X	grf03	sin	vrsta podataka	5	1
N2X	grf03	sin	vrsta reči	30	1
N2X	grf03	sin	vrsta slova	4	1
N2X	grf03	sin	zahtev korisnika	6	1
AXN	grf01	sin	zajednički cilj	6	1
AXN	grf01	sin	zajednički katalog	5	1
AXN	grf01	sin	zajednički rad	10	1
AXN2X	grf05	sin	zajedničko korišćenje resursa	9	1
N4X	grf04a	sin	zakon o bibliotekama	7	1
AXN	grf01	sin	zakonska regulativa	10	1
N4X	grf04b	sin	zaštita bibliotečkog materijala	4	1
N4X	grf04b	sin	zaštita spomenika kulture	8	1
AXN	grf01	sin	zavičajna bibliografija	44	1
AXN	grf01	sin	zavičajna zbirka	15	1
AXN	grf01	sin	zavičajni fond	15	1
N2X	grf03	sin	zbirka knjiga	6	1
AXN	grf01	sin	žensko ime	12	1
N2X	grf03	sin	zgrada biblioteke	6	1

Прилог 2а – Вишечлане леме које су се већ налазиле у речнику DELAC

Образак	graf	broj	lema	FrekGrafLema	ImaUDelacu
AXN	grf01	sin	alfabetски индекс	4	2
AXN	grf01	sin	alfabetски редослед	4	2
2XN	grf02	sin	audio-zapis	7	2
AXN	grf01	sin	autorsko pravo	65	2
N2X	grf03	sin	baza podataka	177	2
AXN2X	grf05	sin	bibliografska baza podataka	7	1
AXN	grf01	sin	bibliografska klasifikacija	4	1
AXN	grf01	sin	bibliografska referenca	7	1
AXN	grf01	sin	citatna analiza	14	2
AXN	grf01	sin	decimalna klasifikacija	22	1
N2X	grf03	sin	diseminacija informacija	13	2
N2X	grf03	sin	eksplozija informacija	4	2
AXN	grf01	plu	elektronske informacije	9	1
AXN	grf01	sin	elektronski katalog	20	2
AXN	grf01	sin	elektronski oblik	12	2
AXN	grf01	sin	elektronski tekst	5	1
AXN	grf01	sin	geografsko ime	7	2
AXN	grf01	sin	govorni jezik	6	2
AXN	grf01	sin	informaciona delatnost	24	2
AXN	grf01	sin	informaciona tehnologija	28	2
AXN	grf01	plu	informacione potrebe	7	1
AXN	grf01	sin	informacioni prostor	4	2
AXN	grf01	sin	informacioni sistem	76	2
N2X	grf03	sin	izvor informacija	12	1
N2X	grf03	sin	izvor podataka	13	1
N2X	grf03	sin	jezik originala	5	1
N2X	grf03	sin	klasifikacija dokumenata	5	1
AXN	grf01	sin	klasifikacioni sistem	9	1
AXN	grf01	sin	lisni katalog	12	2
AXN	grf01	sin	naslovna strana	43	2
N2X	grf03	sin	nosilac informacija	5	2
N2X	grf03	sin	obrada podataka	22	2
AXN	grf01	sin	pravno lice	9	2
AXN	grf01	sin	predmetna klasifikacija	5	1
N2X	grf03	sin	pristup informacijama	13	1
AXN	grf01	sin	programski jezik	42	1
AXN	grf01	sin	programski paket	4	2
N2X	grf03	sin	protok informacija	10	2
N2X	grf03	sin	provera pravopisa	5	2
AXN	grf01	sin	radna jedinica	8	2
AXN	grf01	sin	radna stanica	8	2
AXN	grf01	sin	radno vreme	34	1
N2X	grf03	sin	razmena informacija	7	2
AXN	grf01	sin	redni broj	12	2

AXN	grf01	sin	referensna literatura	11	1
AXN	grf01	sin	referensna zbirka	16	1
AXN	grf01	sin	sistematski katalog	6	1
AXN	grf01	sin	softverska piraterija	22	2
AXN	grf01	sin	softverski alat	7	2
AXN	grf01	sin	upravni odbor	9	2
NXN	grf06	sin	veb lokacija	22	1
NXN	grf06	sin	veb prezentacija	4	1
NXN	grf06	sin	veb stranica	9	1

Прилог 26 – Вишечлане леме екстраховане у множини (облик у јединици прихватљив за термилошки речник)

Образас	граф	број	лема	ФрекГрафЛема
AXN	grf01	plu	akademski izdavači	4
AXN	grf01	plu	aplikativni programi	4
AXN	grf01	plu	bibliografski elementi	6
AXN	grf01	plu	bibliografski izvori	9
AXN	grf01	plu	bibliografski izvori	9
AXN	grf01	plu	bibliografski opisi	8
AXN	grf01	plu	bibliografski principi	10
AXN	grf01	plu	bibliografski servisi	4
AXN	grf01	plu	bibliografski zapisi	34
AXN	grf01	plu	bibliotečki fondovi	27
AXN	grf01	plu	bibliotečki kadrovi	4
AXN	grf01	plu	bibliotečki katalozi	9
AXN	grf01	plu	bibliotečki radnici	42
AXN	grf01	plu	bibliotečki sistemi	15
2XAXN	grf10	plu	bibliotečko-informacioni sistemi	8
AXN	grf01	plu	centralni katalozi	8
AXAXN	grf07	plu	ćirilске рукописне књиге	4
AXN	grf01	plu	citatni indeksi	6
AXN	grf01	plu	dnevni listovi	6
AXN	grf01	plu	elektronske publikacije	9
AXN	grf01	plu	elektronski časopisi	8
AXN	grf01	plu	elektronski dokumenti	12
AXN	grf01	plu	elektronski izvori	37
AXN	grf01	plu	elektronski rečnici	15
AXN	grf01	plu	elektronski resursi	8
AXN	grf01	plu	elektronski servisi	7
AXN	grf01	plu	elektronski tekstovi	8
AXN	grf01	plu	imenski prostori	7
AXN	grf01	plu	informacione usluge	4
AXN	grf01	plu	informacioni centri	11
AXN	grf01	plu	informacioni izvori	29
AXN	grf01	plu	informacioni resursi	5
AXN	grf01	plu	informacioni servisi	4
AXN	grf01	plu	informacioni sistemi	27
AXN	grf01	plu	informacioni stručnjaci	4
AXN	grf01	plu	interesne grupe	4
AXN	grf01	plu	istorijski dokumenti	5
AXN	grf01	plu	istorijski izvori	8
AXN	grf01	plu	istraživačke biblioteke	10
AXN	grf01	plu	istraživačke institucije	4
AXN	grf01	plu	istraživačke metode	5
AXN	grf01	plu	istraživačke organizacije	5
AXN	grf01	plu	istraživački projekti	14

AXN	grf01	plu	izdavački katalozi	4
AXN	grf01	plu	kataloški listići	29
AXN	grf01	plu	kataloški zapisi	4
AXN	grf01	plu	knjižarski katalozi	8
AXN	grf01	plu	knjižni fondovi	4
N2X	grf03	sin	kodiranje tekstova	4
AXN	grf01	plu	leksičke jedinice	9
AXN	grf01	plu	lisni katalozi	5
AXN	grf01	plu	međunarodni standardi	35
AXN	grf01	plu	mrežni servisi	8
AXN	grf01	plu	naslovni listovi	4
AXN	grf01	plu	nastavni metodi	4
AXN	grf01	plu	naučni časopisi	14
AXN	grf01	plu	naučni dokumenti	8
2XAXN	grf10	plu	naučno-tehničke biblioteke	4
AXN	grf01	plu	novinski članci	10
AXN	grf01	plu	novinski tekstovi	8
AXN	grf01	plu	numerički pokazatelji	5
N2X	grf03	sin	obrada tekstova	4
AXN	grf01	plu	optički diskovi	7
AXN	grf01	plu	optički mediji	6
N4X	grf04a	plu	podaci o broju	4
N4X	grf04a	plu	podaci o formatu	5
N4X	grf04a	plu	podaci o korišćenju	5
AXN	grf01	plu	predmetne oblasti	6
AXN	grf01	plu	predmetni specijalisti	6
N2X	grf03	sin	promocija biblioteka	7
AXN	grf01	plu	računarski alati	4
AXN	grf01	plu	računarski programi	8
AXN	grf01	plu	računarski resursi	4
AXN	grf01	plu	radni procesi	15
AXN	grf01	plu	radni zadaci	7
AXN	grf01	plu	rukopisna dela	5
N4X	grf04b	sin	sadržaj periodičnih publikacija	4
N4X	grf04a	plu	sadržaji na vebu	6
AXN	grf01	plu	sastavni delovi	11
AXN	grf01	plu	semantičke etikete	4
AXN	grf01	plu	semantički domen	4
AXN	grf01	plu	semantički markeri	8
AXN	grf01	plu	semantički odnosi	4
AXN	grf01	plu	softverski alati	8
AXN	grf01	plu	softverski paketi	4
AXN	grf01	plu	softverski proizvodi	4
AXN	grf01	plu	štamarska slova	9
AXN	grf01	plu	tehničke biblioteke	4
AXN	grf01	plu	univerzitetski repozitorijumi	4
AXN	grf01	plu	zavičajni fondovi	20

Прилог 2в – Вишечлане леме екстраховане и у множини и у једини

Образас	граф	број	лема	FrekGraflема	EvalTerm
AXN	grf01	sin	bibliografski opis	87	1
AXN	grf01	sin	bibliografski zapis	8	1
AXN	grf01	sin	bibliotečki fond	33	1
AXN	grf01	sin	bibliotečki kadar	12	1
AXN	grf01	sin	bibliotečki sistem	42	1
2XAXN	grf10	sin	bibliotečko-informacioni sistem	47	1
AXN	grf01	sin	centralni katalog	39	1
AXN	grf01	sin	elektronski rečnik	14	1
AXN	grf01	sin	imenski prostor	6	1
AXN	grf01	sin	informacioni centar	10	1
AXN	grf01	sin	kataloški listić	6	1
AXN	grf01	sin	kataloški zapis	4	1
AXN	grf01	sin	knjižni fond	13	1
N2X	grf03	sin	kodiranje teksta	9	1
AXN	grf01	sin	međunarodni standard	20	1
AXN	grf01	sin	naslovni list	4	1
AXN	grf01	sin	naučni dokument	5	1
AXN	grf01	sin	predmetni specijalista	10	1
AXN	grf01	sin	radni proces	6	1
AXN	grf01	sin	sastavni deo	34	1
AXN	grf01	sin	štamparsko slovo	4	1
AXN	grf01	sin	zavičajni fond	15	1

Прилог 2г – Вишечлане речи и њихови синоними одабрани за термилошки речник

Образас	граф	број	лема	ФрекГрафЛема	синоним
AXN	grf01	sin	bibliografsko opisivanje	22	bibliografski opis
AXN	grf01	sin	bibliotečka asocijacija	5	bibliotečko udruženje
N2X	grf03	sin	biblioteka fakulteta	10	fakultetska biblioteka
N4X	grf04a	sin	biblioteka na aerodromu	4	aerodromska biblioteka
N4X	grf04a	sin	biblioteka na univerzitetu	17	univerzitetska biblioteka
N2X	grf03	sin	biblioteka univerziteta	68	univerzitetska biblioteka
NXN	grf06	sin	biblioteka-celina	23	biblioteka celina
N2X	grf03	sin	delatnost biblioteke	9	bibliotečka delatnost
N4X	grf04a	sin	evaluacija u bibliotekama	4	evaluacija biblioteka
N2X	grf03	sin	fond biblioteke	27	bibliotečki fond
AXN	grf01	sin	hronološki raspored	4	hronološki redosled
AXN	grf01	sin	hronološki sled	5	hronološki redosled
N2X	grf03	sin	knjiga inventara	4	inventarska knjiga
AXN	grf01	sin	naučna metoda	9	naučni metod?
N2X	grf03	sin	oblast nauke	6	naučna oblast
N2X	grf03	sin	osoblje biblioteke	5	bibliotečko osoblje
N4X	grf04a	plu	poslovi u biblioteci	6	bibliotečki poslovi
N2X	grf03	sin	saradnja biblioteka	5	međubibliotečka saradnja
AXN	grf01	sin	sveučilišna knjižnica	12	univerzitetska biblioteka
N2X	grf03	sin	usluga biblioteke	6	bibliotečka usluga
AXN	grf01	sin	zajednički katalog	5	centralni katalog
NXN	grf06	sin	zgrada biblioteka	4	bibliotečka zgrada

Прилог 3а –Вишечлане речи одабране за општи речник српског језика

Obrazac	graf	broj	lema	FrekGrafLema
AXN	grf01	sin	akademsko obrazovanje	5
AXN	grf01	plu	balkanske zemlje	5
AXN	grf01	plu	balkanski jezici	8
AXN	grf01	sin	bolonjska deklaracija	5
AXN	grf01	sin	briselska konvencija	4
AXN	grf01	sin	bugarski jezik	15
AXN	grf01	sin	centralna zgrada	4
N2X	grf03	sin	cilj projekta	6
AXN	grf01	sin	civilno stanovništvo	8
AXN	grf01	plu	diplomske studije	4
AXN	grf01	sin	diplomski program	5
AXN	grf01	sin	diplomski rad	9
AXN	grf01	sin	doktorska disertacija	21
AXN	grf01	sin	doktorska teza	15
AXN	grf01	sin	doktorski rad	10
AXN	grf01	sin	dramsko pozorište	5
AXN	grf01	sin	državna institucija	4
AXN	grf01	sin	državna uprava	10
AXN	grf01	plu	državne granice	6
AXN	grf01	plu	državni organi	12
AXN	grf01	sin	državni univerzitet	8
AXN	grf01	sin	engleski jezik	65
AXN	grf01	sin	etički kod	4
AXN	grf01	plu	evropske zemlje	7
AXN	grf01	plu	evropski jezici	20
AXN	grf01	sin	finansijska podrška	6
AXN	grf01	sin	finansijsko sredstvo	6
AXN	grf01	sin	fizički prostor	5
AXN	grf01	sin	fizičko vaspitanje	5
AXN	grf01	sin	francuska književnost	4
AXN	grf01	sin	geografski lokalitet	4
AXN	grf01	sin	hrvatska književnost	10
AXN	grf01	sin	hrvatski jezik	7
N2X	grf03	sin	istorija književnosti	13
AXN	grf01	sin	istorijska vrednost	5
AXN	grf01	sin	istraživačka zajednica	4
AXN	grf01	plu	istraživačke aktivnosti	4
AXN	grf01	sin	istraživački centar	4
AXN	grf01	sin	istraživački karakter	5
AXN	grf01	sin	italijanski jezik	11
AXN	grf01	sin	jezička barijera	4
AXN	grf01	sin	jezička pripadnost	5
AXN	grf01	sin	jezički princip	4
AXN	grf01	sin	jezički sistem	4
AXN	grf01	sin	jezičko nasleđe	5
AXN	grf01	sin	jugoslovenska književnost	24

AXN	grf01	sin	južnoslovenska književnost	5
AXN	grf01	plu	južnoslovenski jezici	8
AXN	grf01	sin	kancelarijsko poslovanje	5
2XAXN	grf10	sin	književno-istorijska studija	5
AXN	grf01	sin	kritički osvrt	4
2XAXN	grf10	sin	kulturno-prosvetna delatnost	7
AXN	grf01	sin	kupovna moć	4
AXN	grf01	sin	latinično slovo	5
AXN	grf01	plu	ljudska prava	4
AXN	grf01	sin	ljudski život	5
AXN	grf01	sin	ljusko biće	4
AXN	grf01	sin	međunarodna standardizacija	4
AXN	grf01	sin	međunarodno pravo	4
N2X	grf03	sin	metod evaluacije	4
N2X	grf03	sin	metodologija istraživanja	5
N2X	grf03	sin	ministarstvo nauke	4
N4X	grf04a	sin	ministarstvo za nauku	6
AXN	grf01	sin	muški pol	6
AXN	grf01	sin	muški rod	9
AXN	grf01	sin	narodna poezija	4
AXN	grf01	sin	narodni govor	6
AXN	grf01	sin	narodni jezik	6
AXN	grf01	sin	narodno stvaralaštvo	4
AXN	grf01	sin	nastavni kadar	4
AXN	grf01	sin	nastavni model	4
AXN	grf01	sin	nastavni plan	8
AXN	grf01	sin	nastavni proces	6
AXN	grf01	sin	nastavni program	14
AXN	grf01	sin	nastavno osoblje	17
AXN	grf01	sin	naučna komunikacija	4
AXN	grf01	sin	naučni eksperiment	4
AXN	grf01	sin	naučni karakter	4
AXN	grf01	sin	naučnoistraživačka delatnost	4
2XAXN	grf10	sin	naučno-istraživački karakter	4
2XAXN	grf10	sin	naučno-istraživački proces	11
2XAXN	grf10	sin	naučno-istraživački rad	20
N4X	grf04a	sin	nauka o književnosti	5
AXN	grf01	sin	obrazovni centar	4
AXN	grf01	sin	obrazovni sistem	4
AXN	grf01	sin	politička situacija	4
AXN	grf01	sin	poljska književnost	4
AXN	grf01	sin	pozorišna umetnost	9
AXN	grf01	sin	pravna zaštita	4
AXN	grf01	sin	radni vek	4
NXN	grf06	sin	rang lista	5
AXN	grf01	sin	razvojni put	6
AXN	grf01	sin	rumunski jezik	9
AXN	grf01	sin	ruska književnost	6

AXN	grf01	sin	ruska literatura	4
AXN	grf01	sin	ruski jezik	6
AXN	grf01	sin	slovački jezik	5
AX	grf01	sin	slovenački jezik	13
AXN	grf01	sin	slovenska književnost	6
AXN	grf01	plu	slovenski jezici	8
AXN	grf01	plu	slovenski narodi	4
AXN	grf01	sin	srpska istorija	5
AXN	grf01	sin	srpska književnost	45
AXN	grf01	sin	srpska kultura	24
AXN	grf01	sin	srpski jezik	205
AXN	grf01	plu	srpski književnici	4
AXN	grf01	plu	srpski manastiri	6
AXN	grf01	sin	srpski narod	63
AXN	grf01	sin	srpski rečnik	5
AXN	grf01	sin	srpskohrvatski jezik	12
AXN	grf01	sin	studentska populacija	4
AXN	grf01	plu	studentski domovi	4
AXN	grf01	sin	studijski program	4
AXN	grf01	sin	studijsko putovanje	4
AXN	grf01	plu	tehnički fakulteti	5
N2X	grf03	sin	teorija književnosti	5
AXN	grf01	sin	umetnička vrednost	5
N2X	grf03	sin	ustanova kulture	4
2XAXN	grf10	sin	vaspitno-obrazovni proces	4
AXN	grf01	sin	visokoškolska institucija	5
AXN	grf01	sin	vremenski rok	5
AXN	grf01	sin	zajednički imenitelj	5
N2X	grf03	sin	zaštita prirode	5
N4X	grf04a	plu	zemlje u razvoju	5
AXN	grf01	sin	ženska osoba	5
AXN	grf01	sin	ženski pol	4
AXN	grf01	sin	ženski rod	7
N2X	grf03	sin	značenje reči	5
N2X	grf03	sin	značenje termina	4

Прилог 36 – Вишечлане речи екстраховане у множини (облик у једнини прихватљив за општи речник)

Образас	graf	broj	lema	FrekGrafLema
AXN	grf01	plu	diplomski radovi	5
AXN	grf01	plu	doktorski radovi	10
AXN	grf01	plu	gramatičke kategorije	4
AXN	grf01	plu	jezičke zajednice	5
AXN	grf01	plu	kritički osvrti	5
AXN	grf01	plu	međunarodni projekti	5
AXN	grf01	plu	merne jedinice	4
AXN	grf01	plu	muška imena	5
AXN	grf01	plu	nastavni materijali	9
AXN	grf01	plu	nastavni planovi	15
AXN	grf01	plu	nastavni programi	20
AXN	grf01	plu	naučna dostignuća	5
AXN	grf01	plu	naučne institucije	6
AXN	grf01	plu	naučne ustanove	7
AXN	grf01	plu	naučni instituti	4
AXN	grf01	plu	naučni radnici	23
AXN	grf01	plu	naučni radovi	15
AXN	grf01	plu	naučni rezultati	4
AXN	grf01	plu	naučnoistraživačke ustanove	5
AXN	grf01	plu	naučnoistraživački instituti	4
AXN	grf01	plu	obrazovni programi	4
AXN	grf01	plu	pojavni vidovi	5
AXN	grf01	plu	političke organizacije	4
AXN	grf01	plu	političke stranke	4
AXN	grf01	plu	pozorišne predstave	9
AXN	grf01	plu	predškolske ustanove	6
AXN	grf01	plu	privredni subjekti	4
AXN	grf01	plu	seminarski radovi	5
AXN	grf01	plu	specijalistički radovi	5
AXN	grf01	plu	televizijske emisije	8
AXN	grf01	plu	univerzitetski profesori	4
AXN	grf01	plu	vremenski okviri	8
AXN	grf01	plu	vremenski periodi	5
AXN	grf01	plu	vremenski razmaci	5
AXN	grf01	plu	zakonski propisi	6

Прилог 4 – Вишечлане речи које су одабране и за термиолошки и за општи речник

Образас	граф	број	лема	ФрекГрафЛема
AXN	grf01	sin	alfabetски принцип	4
2XAXN	grf10	sin	аудио-визуелни материјал	4
AXN	grf01	sin	азбучни ред	8
AXN	grf01	sin	азбучни редослед	7
AXN	grf01	sin	циљна група	5
AXN	grf01	sin	ћирилично писмо	6
AXN	grf01	sin	дискусиона група	5
N2X	grf03	sin	друштво знања	4
AXN	grf01	sin	географска карта	5
AXN	grf01	sin	грађанска ћирилица	4
AXN	grf01	sin	хијерархијска структура	6
AXN	grf01	sin	хронолошки редослед	8
AXN	grf01	sin	информациона писменост	11
N2X	grf03	sin	институција културе	7
AXN	grf01	sin	историјска перспектива	8
AXN	grf01	sin	истраживачки рад	42
AXN	grf01	sin	издавачка кућа	8
AXN	grf01	sin	кадровска политика	6
AXN	grf01	sin	кључна реч	48
N2X	grf03	sin	контрола квалитета	7
AXN	grf01	sin	корисничко име	7
AXN	grf01	sin	латинично писмо	8
AXN	grf01	plu	људски ресурси	6
AXN	grf01	sin	магистарски рад	13
N4X	grf04a	sin	материјал за учење	14
AXN	grf01	sin	међународна конференција	13
AXN	grf01	sin	међународна организација	14
AXN	grf01	sin	међународна сарадња	11
AXN	grf01	sin	међународни карактер	8
AXN	grf01	sin	међународни стандард	20
N2X	grf03	sin	мобилност стручњака	4
AXN	grf01	sin	набавна политика	27
N2X	grf03	sin	нацрт стандарда	9
AXN	grf01	sin	надзорни одбор	12
AXN	grf01	sin	научна баštина	6
AXN	grf01	sin	научна јавност	7
AXN	grf01	sin	научна област	28
AXN	grf01	sin	научни документ	5
AXN	grf01	sin	научни рад	62
AXN	grf01	sin	научни радник	7
AXN	grf01	sin	научно истраживање	21
N2X	grf03	sin	област знања	5
AXN	grf01	sin	образовна институција	6

AXN	grf01	sin	obrazovni proces	17
AXN	grf01	sin	organizaciona jedinica	10
AXN	grf01	sin	organizaciona struktura	11
AXN	grf01	sin	pravni status	7
AXN	grf01	sin	programski savet	8
AXN	grf01	sin	projektni tim	4
AXN	grf01	sin	prosvetna delatnost	7
AXN	grf01	sin	računarska tehnologija	17
AXN	grf01	sin	radna grupa	28
AXN	grf01	sin	radna organizacija	6
AXN	grf01	plu	radne obaveze	5
AXN	grf01	sin	radni proces	6
AXN	grf01	sin	radni prostor	4
AXN	grf01	sin	radno mesto	12
AXN	grf01	sin	radno okruženje	4
AXN	grf01	sin	sastavni deo	34
N2X	grf03	sin	sistem kvaliteta	21
AXN	grf01	sin	slovensko pismo	10
N2X	grf03	sin	spomenik kulture	10
AXN	grf01	sin	statistički metod	5
AXN	grf01	sin	strateški plan	11
N2X	grf03	sin	studija slučaja	5
AXN	grf01	sin	štamparsko slovo	4
AXN	grf01	sin	tehnička dokumentacija	4
AXN	grf01	sin	tehnička podrška	8
AXN	grf01	plu	tehničke mogućnosti	4
AXN	grf01	plu	tehničke osobine	4
AXN	grf01	sin	tehničko osoblje	6
AXN	grf01	sin	tehnološka inovacija	6
AXN	grf01	sin	tehnološka podrška	4
AXN	grf01	sin	tehnološki napredak	10
AXN	grf01	sin	tehnološki razvoj	11
AXN	grf01	sin	terminološki rečnik	6
AXN	grf01	sin	timski rad	8
N4X	grf04a	sin	učenje na daljinu	32
AXN	grf01	sin	umetničko delo	4
AXN	grf01	sin	upravljačka struktura	4
N2X	grf03	sin	upravljanje znanjem	6
NXN	grf06	sin	veb sajt	5
AXN	grf01	sin	veštačka inteligencija	4
AXN	grf01	sin	visokoškolska ustanova	9
AXN	grf01	sin	visokoškolsko obrazovanje	7
AXN	grf01	sin	vremenski period	19
AXN	grf01	sin	vremenski raspon	7
N2X	grf03	sin	vrsta reči	30
AXN	grf01	sin	zajednički cilj	6
AXN	grf01	sin	zajednički rad	10
AXN2X	grf05	sin	zajedničko korišćenje resursa	9

AXN	grf01	sin	zakonska regulative	10
N4X	grf04b	sin	zaštita spomenika kulture	8
AXN	grf01	sin	žensko ime	12

Прилог 5а – Избор конкорданци из дисертације I

metodama, a korišćena je i raspoloživa bibliotečka statistika. {S} Opštom analitičko-sintetičko okviru organizacione strukture Američke bibliotečke asocijacije i mogućnost da ispune obećanja da, mogu da budu putokaz za unapređenje bibliotečke delatnosti. {S} Cilj analize je da se dođe d stručni poslovi poput revizije i otpisa bibliotečke građe, saradnja sa drugim ustanovama kultur jezika i književnosti (60% fonda), kao bibliotečke jedinice potrebne za nastavu drugih predmet tekator-informatorom i traže neku vrstu bibliotečke usluge. {S} U budućnosti će se smanjiti broj celom svetu i za podršku i orijentaciju bibliotečke zajednice. {S} Napisane su da bi pomogle ško predviđeno da je evidentiran celokupan bibliotečki fond koji sadrži građu i izvore različite p o nacionalnim standardima za obavljanje bibliotečko-informacione delatnosti javnih, školskih, s maća i međunarodna dokumenta; - poznaje bibliotečku delatnost (nabavka, obrada, čuvanje i pruža nate o čika Kovi Zmajju?" Učenici iznose biografske podatke koji su im poznati o K. K. Zmajju. {S} oblik zloupotrebe jezika i imaju svoju ciljnu grupu. {S} Reklamama se apeluje na potrošača i on ormacijama i jednostavno prikupljanje i čuvanje podataka. {S} Veoma važnu ulogu u procesu učenja učenicima kako treba naglasiti pojedine delove teksta, insistira na pravilnoj dikciji i akcento učni i video zapisi na bilo kom mediju, elektronske publikacije na fizičkim nosačima i elektron i pratećih naučno-istorijski materijal; elektronski izvori i pristup ka njima, što omogućava uč osnovu raspoloživih izvora (štampanih i elektronskih kataloga) - Ume da koristi sistem Univerza ajmanje 10 - 12 knjiga, u štampanom ili elektronskom obliku, po učeniku, nastavniku i stručnom a mreža telekomunikacija, kao i dodatna finansijska sredstva kako bi se obezbedio poseban softv pitno-obrazovnog rada. {S} Analiziran je fond biblioteke, saradnja i aktivnosti bibliotečkog i n shod bio bi: učenik će upoznati osnovne funkcije biblioteke. {S} Iz ovakve formulacije nije prec avka u predškolskoj ustanovi posećivali gradsku biblioteku i upoznati su sa njenim radom i osno deo fonda, oko 80 %, izostavljena je iz grafičkih prikaza kako bi se video odnos ostalih struka ilustracije, a za starije čak i ime autora. {S} Dvanaesto pitanje otkriva razloge nečita trani -naziv škole, naziv i vrsta rada, ime i prezime predmetnog nastavnika i učenika i mesto. ebi i druge publikacije. {S} Bibliotečko-informaciona delatnost školske biblioteke opisuje struč otrebne za školu i za lične potrebe. {S} Informaciona pismenost razvija se sedamdestih godina pr onog uređenja i propisa iz bibliotečko-informacione delatnosti; pismene provera znanja iz pred e bi mogle biti primenjene na bilo koje informacione potrebe. {S} Upravo školski bibliotekar pod a, prvenstveno na polju komunikacione i informacione tehnologije. {S} Danas je obrazovanje pod s ijuma koji pomažu u proceni efikasnosti informacionih resursa: {S} 1. {S} Tačnost - koliko su inf redmeta u tematske celine; - primenjuje informaciono-komunikacione tehnologije; - primenjuje ra biblioteke deo jedinstvenog bibliotečko-informacionog sistema Republike Srbije i kao takve pred oški čas, čas samostalnog rada učenika, istraživački rad na času, čas testiranja, kviz-čas, čas

je omogućava korišćenje knjiga i drugih izvora informacija, u rasponu od umetničkih do dokument m upravljanja fondom, fizičke zbirke) i kadrovske politike (III odeljak odeljak: {S} Osoblje bib ale serijske publikacije, muzička dela, kartografska građa, katalogi, kalendari, umnoženi umetn , klasifikacija, signiranje); - poznaje klasifikacione sisteme, posebno UDK klasifikaciju; - po jstvu zahtev da im se dostave najnoviji knjižarski katalogi iz Pešte, Beča, Lajpciga i Berlina, sa nastavnicima u primeni informaciono komunikacione tehnologije; - pruža podršku u razvijanju h zakona i istina. {S} Revolucija u info-komunikacionoj tehnologiji ima univerzalni karakter, on program, kao i zahtevi i interesovanja korisnika biblioteke treba da budu primarni prilikom od čitanju i učenju kod dece, kao i navike korišćenja biblioteke tokom čitavog života; • pružanje); - osposobljava učenika za samostalno korišćenje informacija na različitim medijima; - organi izvor odakle su prikupljene). {S}Svako korišćenje podataka u radu treba da bude naznačeno u fu ka u biblioteci, njihovu samostalnost u korišćenju izvora u školskoj biblioteci - prati individ vaspitno-obrazovni proces i da poboljša kvalitet usluga, između ostalog, nabavljajući i neknjiž u procesu edukacije najpre su pronašli mas-mediji, potom audio- vizuelna sredstva i na kraju r a, poštujući zakone i aktuelne domaće i međunarodne standarde iz oblasti prosvete i kulture; - je stručne skupove u vidu akreditovanih međunarodnih konferencija na kojima istaknuti učesnici }U tabelama 1-25 prikazana je struktura monografskih publikacija po okruzima za 2010, 2011. i 2 ima medijska pismenost nije uključena u nastavne programe, poput Austrije i Italije, medijsko o sa spiskovima lektire. {S} Neophodni su nastavni planovi sa promenjenim spisikom lektire, koja j m procesu i isticala se da se savremeni nastavni proces ne može ni zamisliti bez dobro opremlje orijsko utemeljenje bibliotekarstva kao naučne discipline ne može se ni formalno ni sadržinski tematskog korišćenja referensne zbirke, naučne literature i beletristike." Značajan pomak nast a }10) naučno popularnu literaturu za sve naučne oblasti, primerenu deci i mladima; {S}11) beletr S) Školske biblioteke, kao sastavni deo obrazovne institucije, javljaju se u 19. veku zahvaljuj ekara se očekuje da unapređuje vaspitno-obrazovni proces i da poboljša kvalitet usluga, između ašoj zemlji u cilju doprinosa kvalitetu obrazovnog procesa, poslednjih godina česte su reforme za decu i mlade. {S} Glomazna dvostruka organizaciona struktura nije pružala mogućnost da škols ublike Srbije i kao takve predstavljaju organizacione jedinice obrazovne ustanove, s ciljem da e dobijaju dovoljnu autonomiju u okviru organizacione strukture Američke bibliotečke asocijacija zivanja u tekstu. - Pronalazi i izdvaja osnovne informacije iz teksta prema datim kriterijumima i sliku životinje, a na drugoj zapisati osnovne podatke o njoj. {S}Konj Konji mogu spavati i u imu je slika Jovana Jovanovića Zmaja sa osnovnim podacima: {S}Jovan Jovanović Zmaj Rođen je u N a zatim drugi umesto tih reči izgovara osnovno značenje. {S} Razgovaramo kako je pesma zvučala rpski jezik. {S} Izlazi od 1932. godine, periodičnost izlaženja je tromesečna. {S}- Školski čas školska biblioteka može da ponudi radi poboljšanja kvaliteta obrazovanja, a koje obuhvataju in pripremio, kako bi učenicima pomogao u pronalaženju informacija. {S} Zatim treba da, koristeći

ti i razvijanju sistematskog korišćenja [referensne zbirke](#), naučne literature i beletristike.“ Z ljanje literature o temi (monografske i [serijske publikacije](#), članci, enciklopedije, pretraživa oteke i čitalačkih navika podržana je i [studijom slučaja](#).) {S} Dobijeni rezultati mogu da posluže u u prilog autorovim razmišljanjima? {S}Školska biblioteka obezbeđuje elektronske izvore inform šanja statusa i unapređenja delatnosti, [školske biblioteke](#) pored domaćih zakona i standarda pri luga koje se pružaju korisnicima usluga [školskih biblioteka](#). {S}Nakon Drugog svetskog rata osno Švedskoj. {S}Posebnu pažnju posvetio je [školskim bibliotekama](#) u Americi i njihovom razvoju toko obodnih aktivnosti; - planira sekcije u [školskoj biblioteci](#) (bibliotečka, izdavačka, dramska, r rednu nastavu, odnosno na učenje u i sa [školskom bibliotekom](#), otvoreno je pitanje nepostojanja nika: - ženski 58% - muški 42% {S}2. {S}Školsku biblioteku učenici koriste: - jednom ili dvaput ja koje nude ne samo elektronski, već i [štampani izvori](#). {S} U ovome je razlika između informaci mbinuje karakteristike e- obrazovanja i učenja na daljinu je e-učionica. {S} To je okruženje nas nskih kataloga) - Ume da koristi sistem [Univerzalne decimalne klasifikacije](#) kako bi pronašao pu školskih, specijalnih, visokoškolskih i [univerzitetstskih biblioteka](#) i biblioteka naučno-istraživ e da organizuje školsku, visokoškolsku, [univerzitetstsku biblioteku](#), odnosno biblioteku naučnoist : {S} Bibliotekarstvo UDK 027.8 {S}1. {S} [Uvodne napomene](#) Školske biblioteke predstavljaju poseba orisnika putem četa. {S} Korisnici putem [web pretraživača](#) uspostavljaju komunikaciju sa bibliote ionalne biblioteke, preko medijateke do [virtuelnih biblioteka](#) nudeći učenicima informacije u el donošenja standarda za srednje škole i [visokoškolske ustanove](#)28, zahvaljujući naporima Naciona njem obuke; {S}2) akreditovani programi [visokoškolskih ustanova](#) kao oblika celoživotnog učenja; Analiza fondova školskih biblioteka. {S} [Zajedničkim radom](#) školski bibliotekar i nastavnik jezik tupke i ovladaju misaonim procesima. {S} [Zajedničkim radom](#) postaviće intelektualne zahteve tako gradi atmosferu međusobnog poverenja u [zajedničkom radu](#) u interesu učenika; - aktivno i konstr čke kompetencije - Poznajе i primenjuje [zakonsku regulativu](#) u oblasti bibliotekarstva i relevant vne kompetencije - Poznajе i primenjuje [zakonsku regulativu](#) i relevantna dokumenta u obrazovanj ka. {S} Sad je sajt biblioteke zamena za [zgradu biblioteke](#). {S}Otvoren obrazovni sistem prema in

1.1 Прилог 56 – Избор конкорданци из дисертације II

tantom - teškom finansijskom situacijom [akademskih biblioteka](#) Boukacem, Sch?pfel, . {S}Do poras ijedna biblioteka nije uradila detaljnu analizu korišćenja otvorenog pristupa u istraživačkoj i potrebu tehničkih sredstava u korišćenju [autorskih dela](#). {S} I ovde se autorskopravna zaštita pro ročito posle . godine, u pokušaju da se [autorsko delo](#) zaštiti i u novonastalim interaktivnim ko je konvencionalnim putem preko podataka [Banke podataka](#) časopisa, licencirani materijal i otvore jučak da operacionalizacija univerzalne [bibliografske kontrole](#) mora da bude nužno povezana sa u domaćih časopisa Tabela : {S} Održavanje [bibliografskih baza podataka](#) u Univerzitetској библиоте Biomedicina Jugoslavica do . godine oko [bibliografskih jedinica](#), što je utemeljilo osnove za ko aze podataka kakve su PASCAL sa miliona [bibliografskih podataka](#), iz oblasti nauke, tehnologije macione tehnologije, kako bi se miliona [bibliografskih referenci](#) učinilo dostupnijim a samo pre a pozajmljuje će ispuniti formular [bibliografskim podacima](#) kojima raspolaze i ako je potre h disertacija u punom tekstu i pristupu [bibliografskim podacima](#) iz preko časopisa i oko miliona neke veće biblioteke ili informacionog [bibliografskog centra](#), koji raspolazu opštim katalogom; , uz aspekte jedinstvene klasifikacije, [bibliografskog opisa](#) i kataloga, krajem XIX veka pisali nceptiji paralelnoj onoj o Univerzalnoj [bibliografskoj kontroli](#). {S}Već, osnovana je i Međunar pnosti javnosti zaštićenih dela. {S} Kao [bibliotečki servis](#) sa velikom tradicijom, ona je do sko ima i studijama, ali i čuvanju i zameni [bibliotečkih fondova](#) i međubibliotečkoj pozajmici. {S} U j pozajmici. {S} U zavisnosti od razvoja [bibliotečkih servisa](#) i informacionih tehnologija, javlj og sistema informisanja, na prvom mestu [bibliotečkim informacionim sistemom](#) BIS Strategija tehn bnovile saradnju, razvijajući autonomne [bibliotečko-informacione sisteme](#) na platformi ovog orga predložених активности Bibliografskog i [bibliotečkog centra](#) u promovisanju međunarodnog protoka icu i Odbora za međunarodnu pozajmicu i [centralne kataloge](#), da bi . godine prerastao u Međunaro . {S} U mnogim zemljama postoje, takođe, [centralni katalogi](#) za određene oblasti, najčešće za nau servisa su bibliografsko indeksiranje, [citatni indeksi](#) Chemical Abstracts, Science Citation In ksa, kao svojevrsne dopune međunarodnim [citatnim indeksima](#) Thompson - ISI. {S} SCIndeks održava ičko korišćenje izvora, kao najviši i u [današnje vreme](#) podrazumevajući vid saradnje među biblio držaja, aktivnosti i uloga biblioteka u [današnjem vremenu](#) Chartron, Boukacem, . {S} Do izvesnog plementacijom direktive EZ iz . godine, [digitalne kopije](#) su legalne samo za privatnu ili naučnu uka, Fondacija Andrew Mellon., a onda i [digitalni arhiv](#) interdisciplinarne naučne literature JS i neophodnost prilagodavanja fizičkih i [digitalnih biblioteka](#) savremenim trendovima, ovi projek zakona ne zabranjuje eksplicitno slanje [digitalnih kopija](#), neke biblioteke ih šalju, neke se od i izdavača otvorenog pristupa slobodnog [digitalnog arhiva](#) časopisa za biomedicinske nauke PubMe nosti, potenciran kontinualnim razvojem [direktnog pristupa](#) dokumentu, u smislu da između korisn ge strane, mogućnost jednostavne i brze [diseminacije informacija](#) i distribucije materijala, kak

kladu sa svojim fondovima, organizovale [dokumentacione centre](#), koji bi, dalje, razmenjivali kak treba izvršiti napore da se unaprijedi [dostupnost publikacija](#) povećanjem raspoloživosti i brzi ćanja autorske naknade, dozvoljeno je u [državnim organima](#), obrazovnim institucijama i javnim bi e biblioteke - Kraljevskom bibliotekom, [Državnom bibliotekom](#) u Aarhusu i Danskom tehničkom bibl odernim tehničkim metodama pošta, faks, [elektronska pošta](#) dozvoljeno, ali da se izdavaču na rač orijuma članci se pohranjuju u otvorene [elektronske arhive](#) koji odgovaraju standardima Inicijata ko i zbog razloga ekonomske prirode. {S} [Elektronske informacije](#) su najčešće znatno skuplje od š u, bar u ovom periodu. {S} U isto vreme, [elektronske knjige](#) su postale resurs kome se korisnici sada mnogobrojne: - [podešavanje servera elektronske pošte](#) za blokiranje sadržaja osetljive prir ma, [integriše u potpunosti sve dostupne elektronske resurse](#) i nosi sa sobom sve forme digitalno kakav je na primer [PubMed Central PMC](#), [elektronski arhiv](#) za biomedicinsku literaturu, izdavači {S} Zanimljivo je pomenuti da su se prvi [elektronski časopisi](#) s početka devedesetih godina pojav ćnosti. {S} U pitanju su štampane forme, [elektronski izvori](#), naučna literatura, komercijalna lit itetskoj biblioteci "Svetozar Marković" [Elektronski katalog UBSM](#) Katalog doktorskih disertacija ioteke i SNTIJ uzajamna katalogizacija, [elektronski katalozi](#), baze podataka, internet Uključiva ta "novog prostora" izvori informacija, [elektronski resursi](#), autorska prava, licenciranje i sli iblioteka, jeste i unapređenje pristupa [elektronskim informacijama](#) KoBSON,. {S} Konzorcijum je ć časopisa iz oblasti društvenih nauka u [elektronskom izdavaštvu](#) Walford, Walford, izdavaju pi ed klasične pozajmice korisnici služe i [elektronskom knjigom](#), u najvećem procentu za pristup si korišćenje, pohranjene u odgovarajućem [elektronskom obliku](#), u najmanje jednom on lajn arhivu s dukciji brojnih nepublikovanih kataloga [evropskih biblioteka](#) - osnivanju i razvijanju nacionaln m bibliotekama, bibliotekama akademija, [fakultetskim bibliotekama](#) i specijalnim bibliotekama i rukturu softvera, kao i vidove pravne i [finansijske podrške](#) autorima. {S} Povodom godina od osni re nego što se uputi na viši nivo. {S} [Hijerarhijske strukture](#) su retko zastupljene u celosti , sve dok se, kao osnovni uslov, navodi [ime autora](#). {S} Pored autorstva, "CC'" licence sadrže i likom navođenja izvora, prednost, [pored imena autora](#), daje godini koja se navodi u okruglim zag e podataka integrišu u dokumentacione i [informacione resurse](#) biblioteke. {S} National Library of blioteke do biblioteke. {S} Iako su nove [informacione tehnologije](#) značajno pomogle razvoju delat S} Istraživanja Percepcije biblioteke i [informacioni resursi](#) OCLC, Istraživači i servisi za ot e... {S} Veb strane izdavača Sage .{S} [Informacioni servisi](#) o tekućim istraživanjima Community ih aspekata autorskog i srodnih prava u [informatičkom društvu](#) doslovno ponavlja odredbu o ekskl no da se u većini zemalja oni odnose na [inostranu literaturu](#), kao i činjenica da su zastupljeni rosleđenih međubibliotečkoj pozajmici i [isporuci dokumenata](#) od . godine odgovoreno je iz otvore u svetskoj međubibliotečkoj pozajmici i [isporuci dokumenata](#) takođe, o čemu je dosta pisano Bouk oseđuje, fosil velikog naučnog značaja, [istorijska dokumenta](#), umetničko delo, značajni su za mn izvori informacija i elektronski resursi [Izvori informacija](#) se posmatraju kao najvažnija mesta, nje umnožavanja od strane biblioteke sa [javnim pristupom](#), obrazovnih ustanova, muzeja ili arhiv

opiji ili originalnom dokumentu. {S}RIN korisnici biblioteke nude servise za pronalaženje informacije: uslovno razgraničavanje Iz ugla korisničkih potreba sprovedeno je detaljno istraživanje u poslanu kopiju a cena stoji u vezi sa korisničkim grupama samog servisa. {S}Kada dostavlja li ih resursa, s jedne strane i zajedničko korišćenje izvora, kao najviši i u današnje vreme podri u prilog najveće promene u odnosu na korišćenje resursa same biblioteke. {S}Uspostavljena je tupa i korišćenja resursima -, pristupa krajnjih korisnika -, trajnog pristupa -, naplaćivanja stavku osvajanja univerzalne dostupnosti ljudskog znanja. {S}Svako ima pravo na slobodu mišljenja up definisan je kao „sveobuhvatan izvor ljudskog znanja i kulturne baštine koje je potvrdila na rbije, Katalog doktorskih disertacija i magistarskih radova odbranih na Univerzitetu u Beogr BIBLIOTEČKA POZAJMICA U SRBIJI Klasična međubibliotečka pozajmica u Srbiji - Uvod Sistem u okviru -, trajnog pristupa -, naplaćivanja -, međubibliotečke pozajmice -, predavanja i učenja -. {S}emaju dovoljan budžet. {S}Kada je reč o međubibliotečkoj pozajmici i izdavačima, danas postoji Potkomisija za pravna pitanja u vezi sa međubibliotečkom pozajmicom predlaže i fakultativne klase zahteva primetnog u obe pozajmice, međubibliotečku pozajmicu sa inostranstvom u ovom periodu {S}U društvenim i humanističkim naukama monografska publikacija ostaje dominantan izvor informacija i humanističkih nauka, u kojima se monografska publikacija i dalje pokazuje kao dominantan jmjice, regionalni model odnosio se i na nabavnu politiku, kao i na politiku vođenja zajedničkih sa potražnjom, što se odnosi ne samo na nabavnu politiku, nego i na politiku međubibliotečke posu štampane forme, elektronski izvori, naučna literatura, komercijalna literatura, siva literatura ja i kulturne baštine koje je potvrdila naučna zajednica" Berlinska deklaracija, , a na održiv ima Gillet, str-. {S}Servis Nacionalne naučne biblioteke Instituta za naučne i tehničke informacije avaju vezu između opadanja pretplate na naučne časopise i isporuke fotokopiranog materijala, oc i tokom više od četiri decenije, za sve naučne discipline, i za obe vrste zahteva - primljene i jmjica sa inostranstvom preko centralnih naučnih biblioteka. {S}Izveštaji o radu Univerzitetske ikcija principa i prakse, čak i u svetu naučnih biblioteka? {S} Da li je i na temeljima duhovnosti z knjiga i članaka, kupovini članaka iz naučnih časopisa, pristupu i preuzimanju digitalizovani nformacija i dokumenata, kao i nosilaca naučnih informacija i modela komunikacije u nauci uopšt žavanje je da omogući prenos podataka u računarskoj mreži između dva ili više lica preko posred e najveći nacionalni centri za naučne i tehničke informacije The British Library, INIST, CISTI, a u naučna istraživanja, posebno naučno-tehničke informacije. {S} Pored značaja po sebi, pristup

Биографија аутора

Александра (Јовановић) Тртовац рођена је 4. јула 1974. године у Ваљеву. Дипломирала је 2000. године на Катедри за библиотекарство и информатику Филолошког факултета Универзитета у Београду са просечном оценом 9,47 (девет и 47/100).

Од 2001. године запослена је у Универзитетској библиотеци „Светозар Марковић“ у Београду. Звање вишег библиотекара стекла је 2011. године. Тренутно ради на пословима главног редактора електронског каталога у библиотечком систему COBISS што подразумева контролу квалитета записа у библиографској бази података, израду и редакцију записа за нормативну датотеку, образовање и стручно усавршавање кадрова, те организацију и координацију стручне праксе студената Катедре за библиотекарство и информатику. Такође, ради на изради записа и додели метаподатака за дигиталне објекте у оквиру дигиталних збирки које креира, организује и одржава Универзитетска библиотека „Светозар Марковић“.

Поседује лиценцу за рад у свим сегментима система COBISS и предавач је на курсевима за израду записа за серијске публикације у истом систему. Члан је републичке Комисије за доделу лиценци за рад у узајамној библиографској бази COBIB.SR и републичког Редакторског тима за израду нормативних датотетек.

Члан је и Групе за језичке технологије Универзитета у Београду у оквиру које ради на развоју нових двојезичних лексичких ресурса и врши допуну и надградњу електронских речника српског језика, као и израду термилошке базе података из области библиотекарства и информатике.

Као сарадник учествовала је у четири међународна пројекта из области развоја библиотечких система, дигитализације грађе и креирања дигиталних библиотека, а тренутно је ангажована на међународном пројекту Европског програма за сарадњу у домену научних и технолошких информација – COST. Учесник је и бројних домаћих и међународних конференција и радионица на којима је презентовала више радова.

Уже области њеног интересовања су библиотечка информатика и рачунарска лингвистика, посебно дигиталне библиотеке, проналажење информација и рачунарска лексикографија.

Говори енглески језик, а служи се и француским језиком.

Изјаве о докторској дисертацији

Прилог 1.

Изјава о ауторству

Потписана Александра С. Трговац

број индекса 08092 Д

Изјављујем

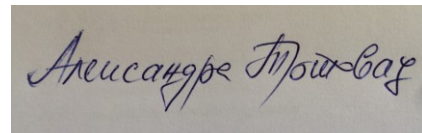
да је докторска дисертација под насловом

Дескриптори метаподатака и дескриптори садржаја у проналажењу информација у дигиталним библиотекама

- резултат сопственог истраживачког рада,
- да предложена дисертација у целини ни у деловима није била предложена за добијање било које дипломе према студијским програмима других високошколских установа,
- да су резултати коректно наведени и
- да нисам кршила ауторска права и користио интелектуалну својину других лица.

У Београду, 08.04.2016

Потпис докторанда



Прилог 2.

**Изјава о истоветности штампане и електронске верзије
докторског рада**

Име и презиме аутора Александра С. Тртовац

Број индекса 08092 Д

Студијски програм Језик – Библиотекарство и информатика

Наслов рада Дескриптори метаподатака и дескриптори садржаја у
проналажењу информација у дигиталним библиотекама

Ментор проф. др Цветана Крстев

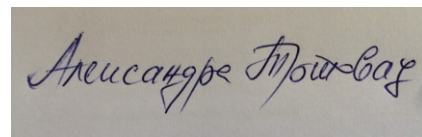
Потписана Александра С. Тртовац

Изјављујем да је штампана верзија мог докторског рада истоветна електронској верзији коју сам предала за објављивање на порталу **Дигиталног репозиторијума Универзитета у Београду**.

Дозвољавам да се објаве моји лични подаци везани за добијање академског звања доктора наука, као што су име и презиме, година и место рођења и датум одбране рада.

Ови лични подаци могу се објавити на мрежним страницама дигиталне библиотеке, у електронском каталогу и у публикацијама Универзитета у Београду.

Потпис докторанда



У Београду, 08.04.2016.

Прилог 3.

Изјава о коришћењу

Овлашћујем Универзитетску библиотеку „Светозар Марковић“ да у Дигитални репозиторијум Универзитета у Београду унесе моју докторску дисертацију под насловом:

Дескриптори метаподатака и дескриптори садржаја у проналажењу информација у дигиталним библиотекама

која је моје ауторско дело.

Дисертацију са свим прилозима предала сам у електронском формату погодном за трајно архивирање.

Моју докторску дисертацију похрањену у Дигитални репозиторијум Универзитета у Београду могу да користе сви који поштују одредбе садржане у одабраном типу лиценце Креативне заједнице (Creative Commons) за коју сам се одлучила.

1. Ауторство

2. Ауторство - некомерцијално

3. Ауторство – некомерцијално – без прераде

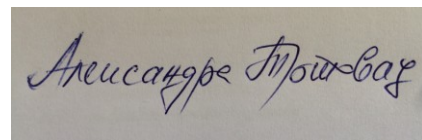
4. Ауторство – некомерцијално – делити под истим условима

5. Ауторство – без прераде

6. Ауторство – делити под истим условима

(Молимо да заокружите само једну од шест понуђених лиценци, кратак опис лиценци дат је на полеђини листа).

Потпис докторанда



У Београду, 08.04.2016.

1. Ауторство - Дозвољаваате умножавање, дистрибуцију и јавно саопштавање дела, и прераде, ако се наведе име аутора на начин одређен од стране аутора или даваоца лиценце, чак и у комерцијалне сврхе. Ово је најслободнија од свих лиценци.
2. Ауторство – некомерцијално. Дозвољаваате умножавање, дистрибуцију и јавно саопштавање дела, и прераде, ако се наведе име аутора на начин одређен од стране аутора или даваоца лиценце. Ова лиценца не дозвољава комерцијалну употребу дела.
3. Ауторство - некомерцијално – без прераде. Дозвољаваате умножавање, дистрибуцију и јавно саопштавање дела, без промена, преобликовања или употребе дела у свом делу, ако се наведе име аутора на начин одређен од стране аутора или даваоца лиценце. Ова лиценца не дозвољава комерцијалну употребу дела. У односу на све остале лиценце, овом лиценцом се ограничава највећи обим права коришћења дела.
4. Ауторство - некомерцијално – делити под истим условима. Дозвољаваате умножавање, дистрибуцију и јавно саопштавање дела, и прераде, ако се наведе име аутора на начин одређен од стране аутора или даваоца лиценце и ако се прерада дистрибуира под истом или сличном лиценцом. Ова лиценца не дозвољава комерцијалну употребу дела и прерада.
5. Ауторство – без прераде. Дозвољаваате умножавање, дистрибуцију и јавно саопштавање дела, без промена, преобликовања или употребе дела у свом делу, ако се наведе име аутора на начин одређен од стране аутора или даваоца лиценце. Ова лиценца дозвољава комерцијалну употребу дела.
6. Ауторство - делити под истим условима. Дозвољаваате умножавање, дистрибуцију и јавно саопштавање дела, и прераде, ако се наведе име аутора на начин одређен од стране аутора или даваоца лиценце и ако се прерада дистрибуира под истом или сличном лиценцом. Ова лиценца дозвољава комерцијалну употребу дела и прерада. Слична је софтверским лиценцама, односно лиценцама отвореног кода.