Импресум

ЗА ИЗДАВАЧА:
проф. др Александар Јерков office@unilib.bg.ac.rs
Универзитетска библиотека „Светозар Марковић“
Филолошки факултет, Универзитет у Београду

ИЗДАВАЧ:
Филолошки факултет, Универзитет у Београду
Универзитетска библиотека „Светозар Марковић“
Заједница библиотека универзитета у Србији

ГЛАВНИ И ОДГОВОРНИ УРЕДНИК:
проф. др Цветана Крстев cvetana@matf.bg.ac.rs
Филолошки факултет, Катедра за библиотекарство и информатику

ОПЕРАТИВНИ УРЕДНИК:
др Александра Тртовац aleksandra@unilib.bg.ac.rs
Универзитетска библиотека „Светозар Марковић“

УРЕДНИК ЕЛЕКТРОНСКОГ ИЗДАЊА:
Јелена Андоновски andonovski@unilib.bg.ac.rs
Универзитетска библиотека „Светозар Марковић“

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

ISSN 1450-9687 (штампано издање)
ISSN 2217-9461 (е-издање)

Београд, год. 17, бр.1, јун 2017.

ИЗРАДА ВЕБ ПОРТАЛА ЧАСОПИСА:
Вукосав Средојевић
Универзитетска библиотека „Светозар Марковић“

ЛЕКТОРИ ЗА ЕНГЛЕСКИ ЈЕЗИК:
Јелена Митровић
Факултет за информатику и математику, Универзитет Пасау, Немачка
др Зоран Ристовић
Основна школа „Мито Игумановић“, Косјерић

ЛЕКТОРИ ЗА СРПСКИ ЈЕЗИК:
др Александра Павловић, Свјетлана Ђелић
Универзитетска библиотека „Светозар Марковић“

ДИЗАЈН И ПРИПРЕМА ЗА ШТАМПУ:
Александар Милошевић, Бранислава Шандрих
и **Инфо***тека* тим

РЕДАКТОР БИБЛИОГРАФИЈЕ И УДК КЛАСИФИКАЦИЈА:
Наташа Дакић
Универзитетска библиотека „Светозар Марковић“

РЕДАКТОР ДОИ БРОЈЕВА:
др Милош Утвић
Филолошки факултет, Универзитет у Београду

РЕДАКЦИЈА ЧАСОПИСА:
Часопис Инфотека
Булевар краља Александра 71, 11000 Београд

+381 11 3370–211
infotheca@unilib.rs

ШТАМПА:
Дунав
Београд

Садржај

Научни радови

Никитас Н. Караниколас, Елени Галиоту
Искуства у изградњи вишедијалекатског речника 7

Василије Милновић, Милена Костић, Матеа Милошевић
Заштита и промоција осетљивог архивског материјала из приватне Библиотеке Лазић 37

Марија Пантић
Отворени образовни ресурси и земље у развоју: један критички осврт 56

Михаило Шкорић
Сврставање појмова на континуалној скали позитивно-негативно према поларитету осећања засновано на емотиконима 72

Стручни радови

Ана Савић
Нормативна контрола у Србији 99

Прикази

Милена Костић, Весна Вуксан, Зоран Бајин
Изложба „Дигитална ћирилица“ 113

Марија Радојичић, Ранка Станковић, Себастијан Каплар
Приказ летње школе о претраживању великих повезаних скупова података заснованом на кључним речима „Keyword search in Big Linked Data“ . . 117

Искуства у изградњи вишедијалекатског речника

УДК 811.14'06'374'282.2.

Никитас Н. Караниколас
nnk@teiath.gr

Елени Галиоту
egali@teiath.gr

Технолошко-образовни институт у Атини
Одсек за информатику

САЖЕТАК: Рад представља искуства стечена током комплетног циклуса изграђивања електронског дијалекатског речника. Иновативност представљеног приступа огледа се у осмишљавању речника који је вишедијалекатски (тројезични), при чему су три дијалекта супротстављена једном циљном језику. Наш приступ отвара могућности које не би биле доступне при раду са три једнојезична дијалекатска речника. Поред наведеног, током успостављања и активирања система искрсавали су веома специфични захтеви за унапређивање система, иницирани од стране корисника, који су се односили посебно на фазу анализе. Стечена искуства и изнађена решења током тежње ка крајњој оптимизацији система у циљу испуњавања свих захтева могу бити корисна за истраживања и пројекте сличне намене.

КЉУЧНЕ РЕЧИ: рачунарска дијалектологија, дијалекатска лексикографија, електронски речници, лексички ресурси, савремени грчки дијалекти, малоазијски грчки.

РАД ПРИМЉЕН: 8. фебруар 2017.
РАД ПРИХВАЋЕН: 28. април 2017.

Превод с енглеског:
Зоран Ристовић

1. Увод

У претходном раду (Karanikolas et. al., 2013), представљен је принцип изградње и начин примене мултимедијалног електронског речника три грчка дијалекта у Малој Азији (понтског, кападочанског и

ајваличког). Представљен је лингвистички и лексикографски приступ, као и принципи изградње макро и микроструктуре речника. Приказан је и концептуални модел вишедијалекатског речника и одговарајућа релациона шема. Према наведеним параметрима осмишљен је и имплементиран систем који садржи леме и релевантне лексикографске информације из три малоазијска грчка дијалекта.

Међутим, током развојног циклуса система, уз помоћ високо квалификованих корисника, искрсавали су веома специфични захтеви за побољшања који су омогућили да постигнемо крајњи циљ, а то је одличан систем. У овом раду приказујемо искуства стечена током изградње система и представљамо унапређени дизајн вишедијалекатског речника и његове додатне могућности. Сматрамо да наш систем може да се примени и на друге грчке дијалекте, а побољшани дизајн може да послужи као база за вишедијалекатске речнике других језика. Систем се може користити за вишедијалекатске речнике других језика ако се обезбеди повезивање са одговарајућим виртуелним тастатурама.

Редослед излагања организован је према следећем плану. Одељак 2. образлаже мотивацију за изградњу речника три дијалекта, а рад на речнику представљен је у одељку 3. Одељци 4. и 5. описују почетне захтеве и нацрт који представља одговор на те захтеве. Одељак 6. износи детаље о првој имплементираној верзији система. Одељак 7. представља захтеве за побољшањима и како је на те захтеве одговорено приликом имплементације. Резултат наведених фаза оптимизације је систем кога одликује функционална изузетност – његов нацрт је тема одељка 8. док су закључна разматрања дата у одељку 9.

2. Мотивација

Понтски, кападочански и ајвалички су три грчка дијалекта Мале Азије која нису довољно документована и којима прети изумирање. До сада је мало пажње поклоњено овим дијалектима. Најинтересантнији изузетак представља Пападопулосов историјски речник понтског дијалекта (Papadopoulos, 1958). Кападочански се помиње у неким другим радовима (Thomason, 2001; Thomason и Kaufman, 1988). Постоје и неки глосари малоазијских грчких дијалеката који садрже речи и идиоматске изразе као и њихово значење на стандардном савременом грчком. Међутим, у већини ових глосара леме су изложене на веома несистематичан начин док суштинске информације као што су изговор и употреба недостају. Поред тога постоји и недоследност у навођењу

глагола – неки глаголи наводе се искључиво у прошлом времену а неки у садашњем. Услед наведених околности темељна анализа грчких дијалеката Мале Азије намеће се као витални приоритет да би се добио увид у природу и механизме језичких промена у домену дијалекатских варијација.

Ови и други релевантни социјални фактори мотивисали су нас да покренемо пројекат AMiGre, у оквиру програма THALIS. Акроним пројекта потиче од кровног наслова иницијативе: „Понт, Кападокија, Ајвалик: у потрази за грчким Мале Азије" (енг. Pontus, Cappadocia, Aivali: in search of Asia Minor Greek). Један од резултата пројекта било је осмишљавање и имплементација мултимедијалног вишедијалекатског речника три грчка дијалекта Мале Азије (понтског, кападочанског и ајваличког), о којима ће бити речи у даљем излагању.

Дијалекатски речници обично се третирају као једнојезични синхронијски речници. У нашем случају (AMiGre), уместо да креирамо три једнојезична дијалекатска речника, одлучили смо да их посматрамо заједно и конципирамо тројезични речник (три малоазијска дијалекта наспрам стандардног савременог грчког језика). У техничком смислу ово је најинтересантнија мотивација која доноси и занимљиву новину јер се омогућава унакрсно референцирање на релацији лема-лема, било да су у питању леме истог дијалекта или различитих дијалеката. Ово не би било изводљиво да су у питању три једнојезична дијалекатска речника.

3. Релевантни радови

Електронска лексикографија модерног грчког језика се донедавно није бавила стварањем дијалекатских речника. Онлајн речници развијени на Порталу за грчки језик (Online, 2016) обухватају рачунарске верзије Георгакасовог *Грчко-енглеског речника*, Триндафилиадесовог *Речника стандардног савременог грчког речника* и обратни речник који су приредили Анастизијади и Симеониди. Поред тога, Портал пружа могућност приступа рачунарској верзији Кријарасовог *Малог речника средњовековне народне грчке књижевности*. Институт за обраду језика и говора развио је двојезичне онлајн речнике (грчко-енглески, грчко-немачки, грчко-руски, грчко-турски и грчко-арапски). Поменути речници се стално развијају и допуњавају а доступни су путем (ILSP, 2016). Поред тога, развијају се и алати за обраду природних језика намењени подршци лексикографских

апликација. У прилог томе, у раду (Tsalidis et. al., 2010) се говори о инфраструктурним алатима који се користе за кодирање морфолошких, синтаксичких и семантичких информација, као и о алатима за коректуру, као што је провера спеловања, поделе речи на крају ретка, итд. Кад је реч о грчким дијалектима, једини нама познати рачунарски обрађени речник је онлајн лексичка база података за кипарски грчки (Themistocleous, 2012). Окружење овог онлајн речника нуди механизме за напредну претрагу као и за претварање текста у говор ради изговора кипарског грчког.

4. Почетни захтеви

Дијалекатски речници обично се третирају као једнојезични синхронијски речници (Béjoint, 2000; Geeraerts, 1989), услед ограничења у микроструктури (општој организацији леме) (Landau, 2001; Zgusta, 1971). Узимајући у обзир да је наша основна намера била да се осмисли и изгради онлајн речник, његова макроструктура не би подлегала физичким ограничењима која су карактеристична за штампане речнике и они би могли да нуде (виртуелно) „вишеструке макроструктуре“ које би одражавале различите опције претраге које смо били у могућности да иградимо (Burke, 2003). С тога, будући да нису постојала ограничења у макроструктури, определили смо се за изградњу тројезичног речника (три малоазијска грчка дијалекта наспрам стандардног савременог грчког језика) (Xydopoulos и Ralli, 2012), уместо три једнојезична дијалекатска речника. Речник је назван TDGDAM (енг. Tri-Dialectal Greek Dictionary of Asia Minor – тродијалекатски речник малоазијског грчког) и он треба да представља лингвистички компетентан тродијалекатски речник у електронском облику. Један од основних захтева TDGDAM-а био је да корисници могу приступити графичком (заснованом на облицима) представљању сваке леме које им допушта да рукују изговором, значењем, употребом и односом са другим лемама. Уређивање овако представљених лема треба да буде омогућено а за то је било потребно да се користе усвојени скупови карактера. Између осталог, свака лема би требало да садржи податке о дијалекатској зони и извору из којег је екстрахована. Овакав тип речника представља иновацију не само када је у питању грчки језик са својим дијалектима, него и када су у питању међународни стандарди, о чему ће бити више речи у даљем излагању.

ΠΕΔΙΟ (FIELD)	ΛΗΜΜΑΤΙΚΗ ΠΛΗΡΟΦΟΡΙΑ (LEMMA VALUES)
1. ΛΕΞΗ-ΚΕΦΑΛΗ / HEADWORD	ΒΡΟΥΛΟ
2. ΛΕΞΙΚΗ ΚΑΤΗΓΟΡΙΑ (lexical category)	(Ο. ουδ.) (noun, neuter)
3. ΦΩΝΗΤΙΚΟΣ ΤΥΠΟΣ (phonetic type)	['vrulu]
4. ΑΡΧΕΙΟ ΗΧΟΥ ΠΡΟΦΟΡΑΣ (digital record)	WAV
5. ΕΝΑΛΛΑΚΤΙΚΟΙ ΤΥΠΟΙ (alternative types)	Βρόλους ['vrolus] (Παμφ. ΜΙΚΡΟΔΙΑΛΕΚΤΙΚΗ ΠΕΡΙΟΧΗ ΣΥΝΔΕΣΗ) (Pamphylia microdialectal region)
6. ΔΙΑΛΕΚΤΙΚΗ ΠΕΡΙΟΧΗ (dialectal region)	Αΐβαλί (Aivali)
7. ΜΙΚΡΟΔΙΑΛΕΚΤΙΚΗ ΠΕΡΙΟΧΗ (microdialect)	
8. ΜΟΡΦΟΛΟΓΙΚΗ ΔΙΕΡΓΑΣΙΑ (morphological process)	-
9. ΧΡΗΣΤΙΚΟ ΣΗΜΑΔΙ (usage)	1. ΦΥΤΟΛΟΓΙΑ (phytology)
10. ΟΡΙΣΜΟΣ (definition)	Βούρλο (bulrush - reedy plant)
11. ΑΡΧΕΙΟ ΕΠΕΞΗΓΗΜΑΤΙΚΗΣ ΕΙΚΟΝΑΣ	JPG
12. ΠΑΡΑΔΕΙΓΜΑ ΧΡΗΣΗΣ (usage example)	«Έκουψα καμπόσα βρούλα κι πέρασα αρμαθιά τα ψάρια πό πιασα σήμερα» (a dialectal sentence using the lemma)
13. ΜΕΤΑΦΡΑΣΗ ΠΑΡΑΔΕΙΓΜΑΤΟΣ ΣΤΗΝ ΚΝΕ (modern Greek translation)	('Εκοψα μερικά βούρλα και .....) (the same sentence in Modern Greek)
14. ΘΗΣΑΥΡΟΣ (thesaurus)	
15. ΕΤΥΜΟΛΟΓΙΚΗ ΠΛΗΡΟΦΟΡΙΑ (etymology)	[ΕΤΥΜ ελνστ. βροῦλον] (originates from the Hellenistic βροῦλον)
9. ΧΡΗΣΤΙΚΟ ΣΗΜΑΔΙ	2. ΥΠΟΤΙΜΗΤΙΚΟ (pejorative)
10. ΟΡΙΣΜΟΣ	Ανόητος (fatuous)
11. ΑΡΧΕΙΟ ΕΠΕΞΗΓΗΜΑΤΙΚΗΣ ΕΙΚΟΝΑΣ	JPG
12. ΠΑΡΑΔΕΙΓΜΑ ΧΡΗΣΗΣ	«Ντιπ για ντιπ βρούλου τούτου του πιδί». (a dialectal sentence using the lemma)
13. ΜΕΤΑΦΡΑΣΗ ΠΑΡΑΔΕΙΓΜΑΤΟΣ ΣΤΗΝ ΚΝΕ	(Τελείως ανόητο αυτό το παιδί) (the same sentence in Modern Greek)
14. ΘΗΣΑΥΡΟΣ	---

Слика 1. Нацрт структурног приказа леме ВРОУЛО

Узимајући у обзир предвиђени географски и временски обухват, TDGDAM је осмишљен као локални дијалекатски речник несинхронијске природе који би требало да обухвати одреднице из различитих просторних зона и временских периода (Penhallurik, 2009). Како је одлучено на самом почетку, леме које улазе у TDGDAM требало би да долазе, директно или индиректно, из говорног или писаног материјала сваког од заступљених дијалекатских варијетета (Keymeulen, 2010).

У погледу микроструктуре TDGDAM-а, наш циљ је био да укључимо формалне информације о изговору (фонетски облик), граматици (категоријалне и морфолошке информације), пореклу (етимологија), значењу (синонимске и дескриптивне дефиниције), употреби (ознаке теме и регистра) као и да обезбедимо везе ка мултимедијалним изворима (унутрашњим или спољашњим у односу на TDGDAM) да би обогатили семантички и прагматички опис лема (Barbato и Varvaro, 2004; Rys и Keymeulen, 2009; Xydopoulos и Ralli, 2012). Како би се избегле разлике и произвољности у писању у оквиру истог дијалекта (Durkin,

2010; Xydopoulos, 2012), одреднице се не појављују у „полуфонетској“ транскрипцији већ се наводе записане великим словима у ортографском облику. Наиме, ортографски облик записан великим словима одудара од писања у стандардном дијалекту; он није прескриптиван када је у питању ортографија дијалекта и допушта да се у микроструктури појаве било који алтернативни ортографски облици (Markus и Heuberger, 2007; Xydopoulos, 2012). Коначно, аутентични примери употребе сматрани су основним саставним делом одредница које се појављују написане у нестандардном облику, одсликавајући што приближније изговор уз употребу дијакритика, али избегавањем „полуфонетске транскрипције“ (Rys и Keymeulen, 2009).

ΠΕΔΙΟ (FIELD)	**ΛΗΜΜΑΤΙΚΗ ΠΛΗΡΟΦΟΡΙΑ**
1. ΛΕΞΗ-ΚΕΦΑΛΗ / HEADWORD	ΛΙΩΣΤΡΑ
2. ΛΕΞΙΚΗ ΚΑΤΗΓΟΡΙΑ (lexical category)	Ο. Θηλ. (noun, feminine)
3. ΦΩΝΗΤΙΚΟΣ ΤΥΠΟΣ (phonetic type)	[‘ʎostra]
4. ΑΡΧΕΙΟ ΗΧΟΥ ΠΡΟΦΟΡΑΣ (digital record)	αρχείο WAV
5. ΕΝΑΛΛΑΚΤΙΚΟΙ ΤΥΠΟΙ (alternative types)	
6. ΔΙΑΛΕΚΤΙΚΗ ΠΕΡΙΟΧΗ (dialectal region)	Αϊβαλί (Aivali)
7. ΜΙΚΡΟΔΙΑΛΕΚΤΙΚΗ ΠΕΡΙΟΧΗ (microdialectal region)	
8. ΜΟΡΦΟΛΟΓΙΚΗ ΔΙΕΡΓΑΣΙΑ (morphological process)	Παραγωγή (derivation)
9. ΧΡΗΣΤΙΚΟ ΣΗΜΑΔΙ (usage)	1. ΥΠΟΤΙΜΗΤΙΚΟ (pejorative)
10. ΟΡΙΣΜΟΣ (definition)	γυναίκα που περιφέρεται εδώ κι εκεί (woman who strolls)
11. ΑΡΧΕΙΟ ΕΠΕΞΗΓΗΜΑΤΙΚΗΣ ΕΙΚΟΝΑΣ (picture)	-
12. ΠΑΡΑΔΕΙΓΜΑ ΧΡΗΣΗΣ (usage example)	«Ξιπόρτσι πάλ’-η λ’ώστρα» (dialectal phrase using the lemma)
13. ΜΕΤΑΦΡΑΣΗ ΠΑΡΑΔΕΙΓΜΑΤΟΣ ΣΤΗΝ ΚΝΕ (Modern Greek translation)	(Ξεπόρτισσε πάλι η γυρίστρα) (the same in Modern Greek)
14. ΘΗΣΑΥΡΟΣ (thesaurus)	Συν: *αλλουγυρίστρα, σόρτα, τακιού* (Synonymy: *αλλουγυρίστρα, σόρτα, τακιού*)
15. ΕΤΥΜΟΛΟΓΙΚΗ ΠΛΗΡΟΦΟΡΙΑ (etymology)	[<*λιέμι* με -ωστρα ίσως από επιδρ. άλλων θηλυκών σε -ωστρα] (*λιέμι* with affix *ώστρα*)
16. ΔΙΑΠΑΡΑΠΟΜΠΕΣ (see also)	

Слика 2. Нацрт структурног приказа леме ΛΙΩΣΤΡΑ

Кад су у питању могућности унакрсног повезивања ставки TDGDAM-а, дефинисали смо три основне потребе: унакрсно референцирање ка другим одредницама, које су повезане било кроз деривационе процесе или кроз семантичке релације; повезивање лема са еквивалентним значењима из различитих дијалеката; синонимске и антонимске релације.

Слике 1, 2 и 3 представљају нацрт структурних описа са подједнаким бројем лема које би TDGDAM требало да садржи. На основу овог

ΠΕΔΙΟ	ΛΗΜΜΑΤΙΚΗ ΠΛΗΡΟΦΟΡΙΑ
1. ΛΕΞΗ-ΚΕΦΑΛΗ / HEADWORD	ΑΛΛΟΥΓΥΡΙΣΤΡΑ
2. ΛΕΞΙΚΗ ΚΑΤΗΓΟΡΙΑ	(Ον. Θηλ.) (noun, feminine)
3. ΦΩΝΗΤΙΚΟΣ ΤΥΠΟΣ	[aluji'ristra]
4. ΑΡΧΕΙΟ ΗΧΟΥ ΠΡΟΦΟΡΑΣ	WAV
5. ΕΝΑΛΛΑΚΤΙΚΟΙ ΤΥΠΟΙ	---
6. ΔΙΑΛΕΚΤΙΚΗ ΠΕΡΙΟΧΗ	Αϊβαλί (Aivali)
7. ΜΙΚΡΟΔΙΑΛΕΚΤΙΚΗ ΠΕΡΙΟΧΗ	---
8. ΜΟΡΦΟΛΟΓΙΚΗ ΔΙΕΡΓΑΣΙΑ	Σύνθετο (composite)
9. ΧΡΗΣΤΙΚΟ ΣΗΜΑΔΙ	1. ΥΠΟΤΙΜΗΤΙΚΟ (pejorative)
10. ΟΡΙΣΜΟΣ	γυναίκα που περιφέρεται εδώ κι εκεί (woman who strolls)
11. ΑΡΧΕΙΟ ΕΠΕΞΗΓΗΜΑΤΙΚΗΣ ΕΙΚΟΝΑΣ	---
12. ΠΑΡΑΔΕΙΓΜΑ ΧΡΗΣΗΣ	"Ξιπόρτσι πάλ'ι η γ'αλλουγυρίστρα" (dialectal phrase using the lemma)
13. ΜΕΤΑΦΡΑΣΗ ΠΑΡΑΔΕΙΓΜΑΤΟΣ ΣΤΗΝ ΚΝΕ	(Πάλι βγήκε η αλλουγυρίστρα.) (the same in Modern Greek)
14. ΘΗΣΑΥΡΟΣ	ΣΥΝ *λυώστρα, σόρτα* (Synonymy: *λυώστρα, σόρτα*)
9. ΧΡΗΣΤΙΚΟ ΣΗΜΑΔΙ	2. ΙΑΤΡΙΚΗ (medicine)
10. ΟΡΙΣΜΟΣ	πόνος με πρήξιμο γύρω από το νύχι (pain & turgescence around nail)
11. ΑΡΧΕΙΟ ΕΠΕΞΗΓΗΜΑΤΙΚΗΣ ΕΙΚΟΝΑΣ	JPG
12. ΠΑΡΑΔΕΙΓΜΑ ΧΡΗΣΗΣ	"Έχου μνιαν αλλουνυρίστρα στου δαχτύλ'ιμ τσι πουν'εί" (dialectal phrase using the lemma)
13. ΜΕΤΑΦΡΑΣΗ ΠΑΡΑΔΕΙΓΜΑΤΟΣ ΣΤΗΝ ΚΝΕ	(Έχω μια αλλουγυρίστρα στο δάχτυλο και με πονάει.) (the same in Modern Greek)
14. ΘΗΣΑΥΡΟΣ	---
15. ΕΤΥΜΟΛΟΓΙΚΗ ΠΛΗΡΟΦΟΡΙΑ	[ΕΤΥΜ από το ρ. *αλλουγυρίζω*] (from the verb *αλλουγυρίζω*)
16. ΔΙΑΠΑΡΑΠΟΜΠΕΣ	---

Слика 3. Нацрт структурног приказа леме ΑΛΛΟΥΓΥΡΙΣΤΡΑ

и сличних нацрта структурних приказа конципирали смо TDGDAM систем.

Раније коришћени појмови синонимије (Synonym, 2016) и антонимије (Opposite, 2016), као и појмови хомонимије (Homonym, 2016) и полисемије (Polysemy, 2016) које ћемо користити касније јасно су дефинисани. Њихове дефиниције могу се пронаћи на интернету.

5. Нацрт

На основу анализе представљене у одељку о почетним захтевима и на основу нацрта структурних приказа (видети слике 1, 2 и 3) као резултат је добијена следећа структура лема:

– Одредница, дијалекат (дијалекатско подручје), морфолошка информација или процес и етимологија су примарне информације са јединственим вредностима које удружене формирају лему и зависе од ње.

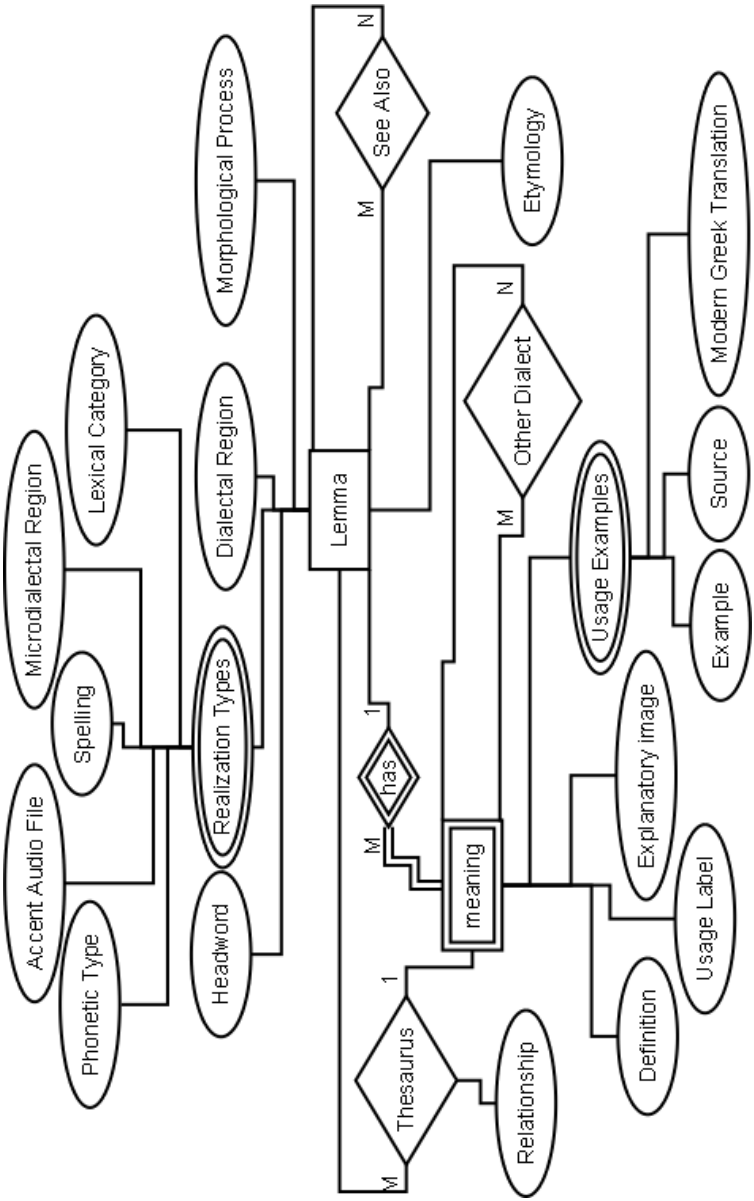
– Свака лема може да има бројне различите реализације а сваку од њих карактерише у извесној мери различита фонетска реализација директно зависна од микродијалекатског подручја са којег потиче – то су специфичне зоне у оквиру ширег дијалекатског подручја у оквиру којег се јавља дата лема.
– Свака лема потенцијално може имати различита значења (појава полисемије), или испољавати хомонимију са другим семантички различитим лемама.
– За свако појединачно значење, различити примери употребе су од суштинског значаја.

Када су у питању релације (између лема и њихових значења) донели смо следеће закључке:

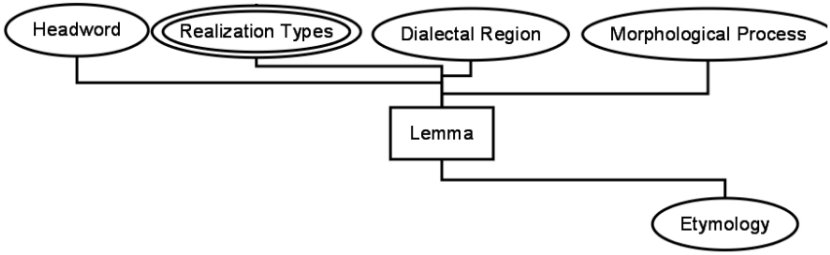
– Унакрсне референце („види и“) могу да буду доступне за повезивање лема које су семантички, прагматички, морфолошки или етимолошки повезане једна са другом.
– *синоними* и *антоними* су две семантичке релације примењиве између различитих лема. Обе наведене релације односе се на значење једне леме у односу на другу, референсану лему. Синонимске и антонимске везе су примењиве на односе између значења леме и леме који потичу из истог дијалекта.
– Постоје значења различитих лема из различитих дијалеката која деле исту дефиницију. Ова релација обележена је са „други дијалекат“. Насупрот другим релацијама, „други дијалекат“ је симетрична релација.

Целокупна идеја (структура лема и њихових релација) строго је дефинисана дијаграмом ентитет-везе на слици 4.

Четири речника података који следе (табеле 1, 2, 3 и 4) објашњавају четири дела (подшеме) целокупног дијаграма ентитет-везе.



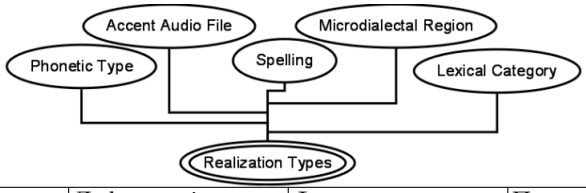
Слика 4. Дијаграм ентитет-везе



Атрибут	Дефиниција	Формат података	пример
Headword (одредница)	Канонски облик речи	Ниска која садржи само велика слова грчког алфабета	ΑΛΛΟΥΓΥΡΙΣΤΡΑ
Etimology (етимологија)	Основне информације о пореклу речи	Ниска записана на грчком са политоналним акцетнима	Από το ρήμα αλλουγυρίζου (од глагола aluji'rizu)
Morphological Process (морфолошки процес)	Различити процеси укључени у формирање речи	Вредност из унапред утврђене листе морфолошких процеса	Σύνθετο (сложеница)
Dialectal region (дијалекатски регион)	Регион или дијалекат у оквиру којих је лема пронађена	Вредност из унапред утврђене листе дијалеката	Αϊβαλί (AIVALI)

Табела 1. Речник података за *Lemma* (лема)

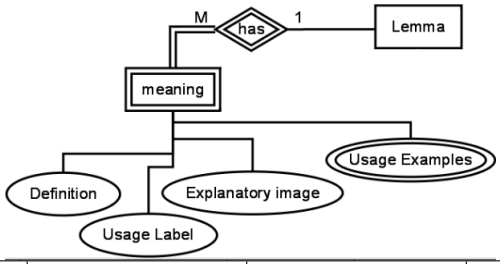
Једна од могућих имплементација концептуалног модела ентитет-везе коришћењем релационе базе података приказана је ралационом шемом на слици 5 која садржи тринаест табела. Међутим, од примарног значаја је седам табела. Осталих шест су помоћне табеле које садрже скупове постојећих доступних вредности за нека поља из осталих седам важних табела. Табеле од примарног значаја су на слици 5 посебно истакнуте подебљавањем ивица и повећавањем фонта њиховог назива. Четири од седам главних табела односе се на релационе еквиваленте главног концептуалног ентитета *Lemma* (лема), слабог ентитета *Meaning* (значење) као и на два сложена атрибута са вишеструким вредностима *Realization Types* (типови реализације) и *Usage Examples* (примери



Атрибут	Дефиниција	Формат података	Пример
Phonetic Type (фонетски тип)	Фонетска транскрипција изговора речи који се изучава	Ниска која садржи карактере Међународног фонетског алфабета (IPA).	aluji'ristra
Accent Audio (аудио акцента)	Аудио датотека која садржи аутентични изговор речи	Ниска која садржи путању датотеке	http://amigre.gr/xyzR1.wav
Spelling (спелинг)	Нестандардна графијска репрезентација изговора у складу са правописним правилима стандардног грчког, комбинована са дијакритицима да би се анотирале фонолошке алтернације	Ниска која садржи слова грчког алфабета и друге дијакритичке симболе (акценте, цртице, заграде и апострофе)	αλλουγυρίστρα
Microdialectical Region (микро-дијалекатски регион)	Име специфичне зоне унутар ширег дијалекатског подручја у коме је реализовани облик леме пронађен	Вредност из унапред прописане листе микро-дијалекатских подручја	
Lexical Category (лексичка категорија)	Врста речи и род	Вредност из унапред прописане листе лексичких категорија	Ουσιαστικό Θηλυκό (именица женског рода)

Табела 2. Речник података за *Realization Types* (типови реализације)

употребе). Три преостале важне табеле су релациони еквиваленти концептуалних релација – *See Also* (види и), *Thesurus* (тезаурус) и *Other Dialect* (други дијалекат).

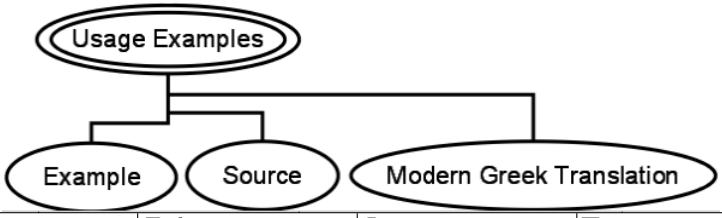


Атрибут	Дефиниција	Формат података	Пример
Definition (дефиниција)	Кратак опис значења леме	Ниска из стандардног савременог грчког	Γυναίκα που περιφέρεται εδώ κι εκεί (“жена која иде наоколо”)
Explanatory Image File (датотека слике за објашњење)	Слика која илуструје значење леме	Ниска која садржи путању датотеке	http://amigre.gr/xyzM1.png
Usage Label (обележје коришћења)	Формална индикација контекста у коме се лема користи (стил, регистар и др.)	Вредност из унапред прописане листе домена	ΥΠΟΤΙΜΗΤΙΚΟ (пежоратив)

Табела 3. Речник података за *Meaning* (значење)

Само табела *Meaning Sets* (скупови значења), као имплементација концептуалне релације *Other Dialect* (други дијалекат), захтева додатно објашњење. Ова релација је симетрична по природи, тј. кад год се за значење одређене леме из неког дијалекта утврди да је еквивалентно значењу неке леме из неког другог дијалекта, онда важи и обрнуто. Управо табела „скупови значења“ и логика саме апликације осигуравају ту симетрију. Преостале две релације, „види и“ и „тезаурус“, нису само по себи симетричне. Ово се очитава у релационој шеми као и у логици саме апликације. Последица тога је да корисник мора да дефинише релацију

у оба смера ако је у неком случају релација „види и“ или „тезаурус“ симетрична.



Атрибут	Дефиниција	Формат података	Пример
Usage example (примери употребе)	Фраза или реченица која служи као пример употребе леме специфичног значења у оригиналном дијалекту	Цео пример (цела ниска) је записан словима грчког алфабета и другим дијакритичким симболима (акценти, цртице, заграде и апострофи)	Ξιπόρτσι πάλ’-η-γ’-αλλουγυρίστρα
Standard Modern Greek Translation (превод на стандарни савремени грчки)	Превод примера употребе на стандардни савремени грчки	Ниска на стандардном савременом грчком	Πάλι βγήκε η αλλουγυρίστρα
Source (извор)	Референца извора из кога је узет пример употребе	Ниска (може да буде књига, URL, итд.)	

Табела 4. Речник података за *Usage Examples* (примери употребе)

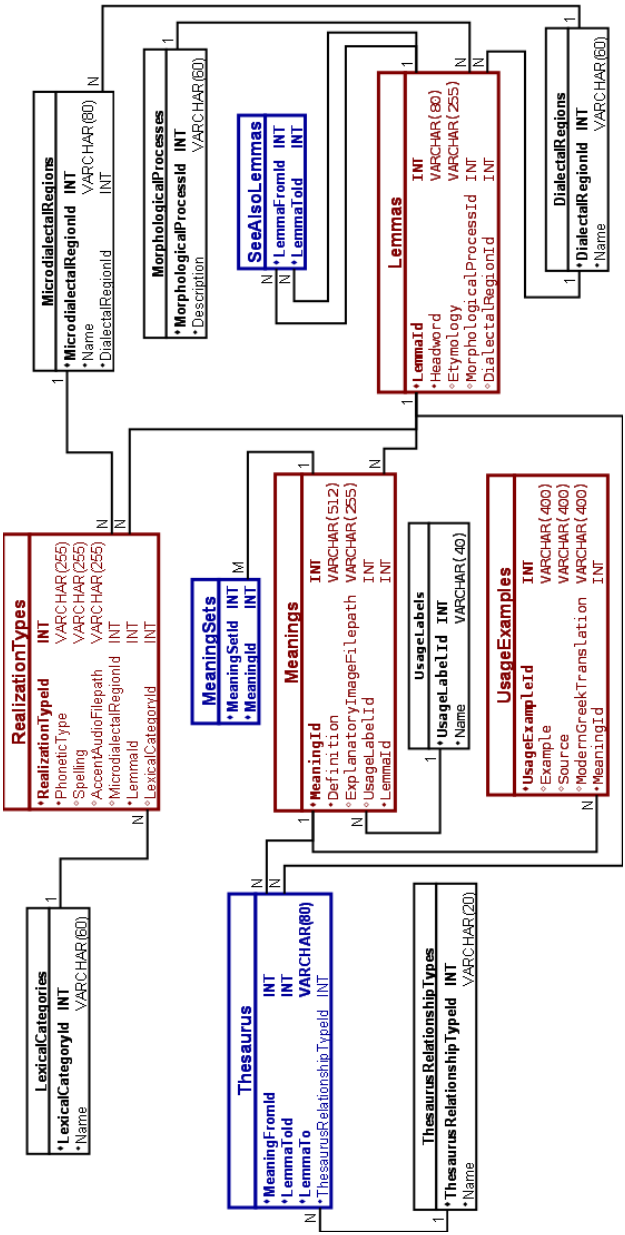
Међународни фонетски алфабет (енг. The International Phonetic Alphabet - IPA) који се користи у табели 2 – подшема *Realization Types* (типови реализације) – је алфабетски систем фонетске нотације који је примарно заснован на латиници. Њега је развила Међународне фонетска асоцијација за стандардизовано представљање звукова говорног језика. IPA стандарде користе лексикографи, студенти и наставници страних језика, лингвисти, патолози говорног језика, певачи, глумци, креатори вештачких језика и преводиоци. Слике 6 и 7 представљају најкорисније IPA листе.

Други приступ фонетској нотацији је SAMPA (Фонетски алфабет заснован на методама процене говора – енг. Speech Assessment Methods Phonetic Alphabet) који представља машински читљив фонетски алфабет. Он је првобитно био развијен у оквиру пројекта ESPRIT 1541, SAM (Методе процене говора – енг. Speech Assessment Methods) у периоду 1987-1989. У почетку је примењен на дански, холандски, енглески, француски и италијански (1989), а касније је примењен и на норвешки и шведски (1992). Потом је примена проширена на грчки, португалски и шпански (1993) да би у скорије време био проширен и на бугарски, естонски, мађарски, пољски и румунски (1996).

6. Имплементација

GUI (графички кориснички интерфејс, енг. Graphical User Interface) верзија система заснива се на два формулара: „главни формулар“ и „формулар за значења“. Слика 8 представља главни формулар леме ΑΛΛΟΥΓΥΡΙΣΤΡΑ. Главни формулар је подељен у три дела. Горњи део пружа информацију о одредници, етимологији, морфолошким процесима и дијалекту. Средишњи део је панел са две картице. Прва картица тог панела служи за приказ и уређивање реализација, док други служи за листе значења лема. Доњи део формулара служи за листу референци типа „види и“. Детаљнији опис средишњег дела формулара дат је на слици 9 која приказује њену другу картицу која даје листу значења за исту лему (представљену на слици 8).

Формулар значења леме позива се преко дугмета за активацију када корисник селектује неку ставку из листе значења која се налази у главном формулару. Слика 10 представља формулар значења који даје нека од значења леме ΑΛΛΟΥΓΥΡΙΣΤΡΑ. Формулар значења је подељен у три дела. Горњи део даје дефиницију значења, а могу бити приложени



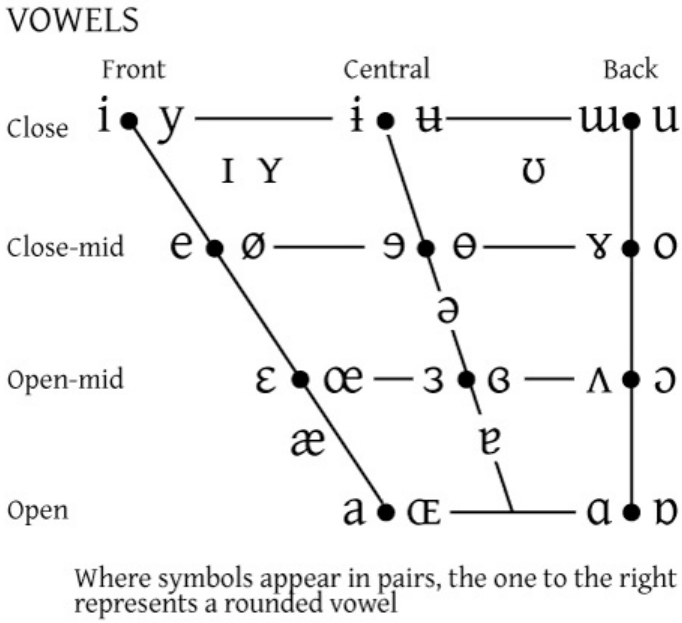
Слика 5. Логичка релациона шема

CONSONANTS (PULMONIC) © 2005 IPA

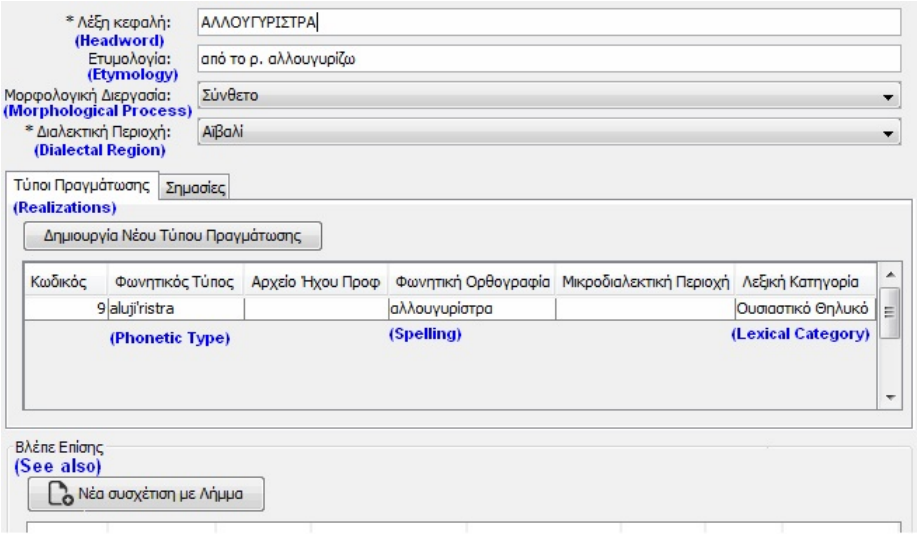
	Bilabial	Labiodental	Dental	Alveolar	Postalveolar	Retroflex	Palatal	Velar	Uvular	Pharyngeal	Glottal
Plosive	p b			t d		ʈ ɖ	c ɟ	k ɡ	q ɢ		ʔ
Nasal	m	ɱ		n		ɳ	ɲ	ŋ	ɴ		
Trill	ʙ			r					ʀ		
Tap or Flap		ⱱ		ɾ		ɽ					
Fricative	ɸ β	f v	θ ð	s z	ʃ ʒ	ʂ ʐ	ç ʝ	x ɣ	χ ʁ	ħ ʕ	h ɦ
Lateral fricative				ɬ ɮ							
Approximant		ʋ		ɹ		ɻ	j	ɰ			
Lateral approximant				l		ɭ	ʎ	ʟ			

Where symbols appear in pairs, the one to the right represents a voiced consonant. Shaded areas denote articulations judged impossible.

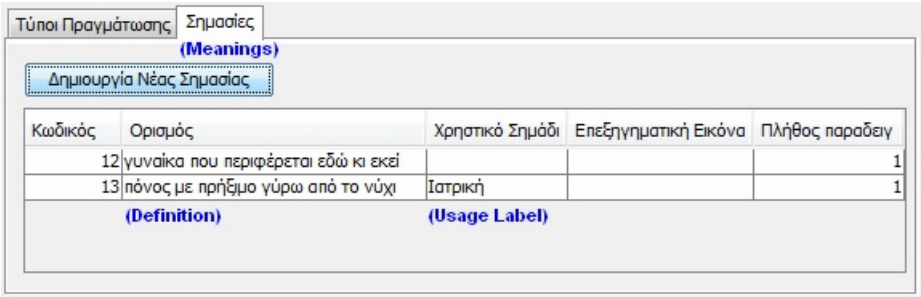
Слика 6. Међународни фонетски алфабет (ревидирана верзија из 2005) – сугласници (пулмонални)



Слика 7. Међународни фонетски алфабет – самогласници

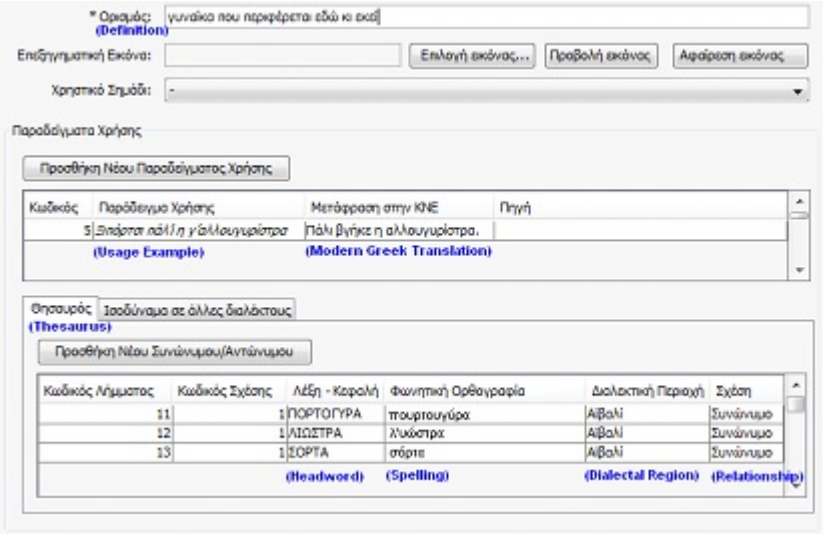


Слика 8. Главни формулар леме ΑΛΛΟΥΓΥΡΙΣΤΡΑ – картица реализација је у првом плану



Слика 9. Главни формулар леме ΑΛΛΟΥΓΥΡΙΣΤΡΑ – картица значења

и слика и додатна информацију о употреби. Средишњи део формулара је панел у коме су наведени примери употребе. Доњи део је реализован као панел са две картице. Прва картица се користи за приказ и уређивање релација синонимије и антонимије (тезаурус), док се друга картица користи за руковање еквивалентима у другим дијалектима.



Слика 10. Формулар значења леме ΑΛΛΟΥΓΥΡΙΣΤΡΑ – нека значења са припадајућим тезаурусом

У складу са оним што је речено у одељцима о анализи (захтевима) и нацрту, реализовани систем нуди кориснику следеће скупове карактера за уређивање одговарајућих поља:

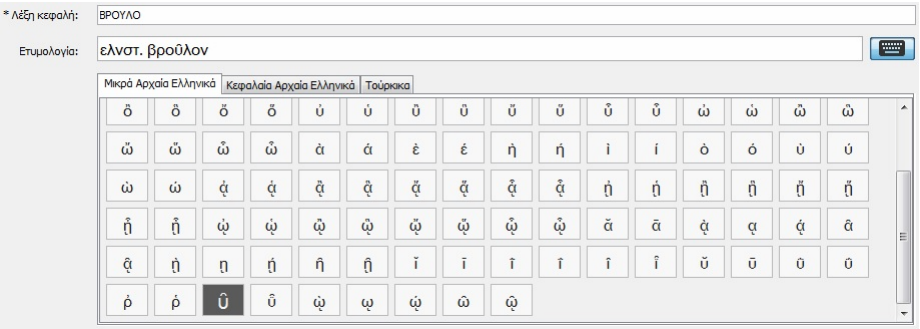
- Етмологија
  - грчки политонални;
  - позајмљени карактери из других алфабета уколико су у питању позајмљенице (нпр. карактери из турског алфабета);
- Фонетски тип
  - IPA;

– Спелинг (фонетска ортографија)
  - савремени грчки;
  - акценти;
  - цртице, заграде и апострофи.

7. Завршна оптимизација

1. Као што је добро познато, корисници најчешће изражавају своје мишљење у завршним фазама развојног циклуса система (*прихватање*, *инсталација*, *развој*) као и током процеса одржавања. И у случају нашег електронског речника било је тако. У нашем случају корисници су изнели гледиште да би атрибут порекла (етимологија) леме требало да се денотира поштујући изворе и конвенције језика из којег и потиче, то јест који је извор одређене леме. Према томе, у случају вишедијалекатског речника атрибут етимологије може да садржи речи из било ког од језика који су имали утицаја на одређени дијалекат. Да би се овоме удовољило понудили смо решење да се обезбеди (у каснијем развоју) виртуална тастатура за сваки језик који је вршио утицај. У случају три дијалекта који су заступљени у оквиру система AMiGre алфабети који су остварили утицај су савремени грчки, старогрчки и турски. Слика 11 представља виртуалну тастатуру са три картице за мала слова старогрчког, велика слова старогрчког и за турски. Они су, заједно са савременим грчким за који је на располагању физичка тастатура омогућавали корисницима да унесу жељену етимологију за сваку појединачну лему без икаквих ограничења.

2. Када су у питању атрибут *Phonetic Type* (фонетски тип) и интонација самогласника постоје две могућности. Прва могућност је навођење акцента пре наглашеног слога. Друга могућност је да се употреби наглашена верзија самогласника. Обе опције су семантички еквивалентне. Међутим, прва опција је лакша за употребу и имплементацију у систему јер нема потребе да се подржавају обе верзије IPA симбола које се користе за самогласнике, са и без акцента. Тако је током иницијалне развојне фазе наш систем допуштао само вертикалну акцентуацију пре слога за обележавање интонираног самогласника. Због тога је на слици 8 *Phonetic Type* обележен са “aluji’ristra”. Након иницијалне развојне



Слика 11. Виртуална тастатура за старогрчки и турски

фазе и током фазе одржавања система указала се потреба да се подрже акцентоване верзије самогласника. Како би се ова потреба задовољила, проширили смо систем тако да обухвати и виртуалне тастатуре за лако уметање IPA симбола, са или без акцената. Слике 12, 13, 14 и 15 представљају виртуалне тастатуре коришћене за уношење вредности атрибута *Phonetic Type*.

Слика 12. Алфабетски (IPA) симболи

Комбинације IPA симбола (слика 12) са неким од дијакритичких симбола (слика 13) творе акцентоване IPA симболе. Слика 16

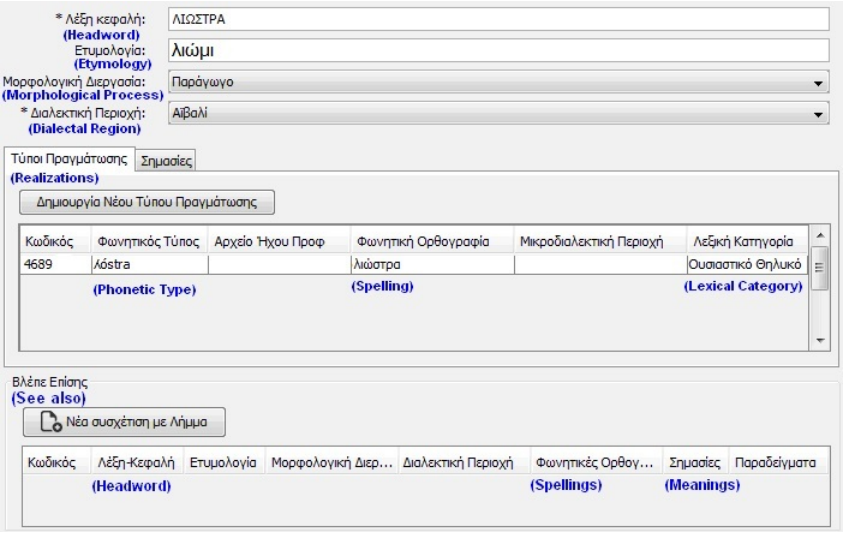
Слика 13. Дијакритички симболи

представља још једну лему - ΛΙΩΣΤΡΑ – која има фонетски тип са акцентованим IPA симболима.

Слика 14. Сложени симболи

Слика 15. Симболи са стрелицама

3. На основу лингвистичке анализе, атрибут *spelling* (спеловање) представља нестандардну графичку презентацију изговора речи према ортографским правилима савременог грчког језика (циљног језика), комбиновану са дијакритицима како би биле забележене све фонолошке алтернације. Сходно томе дошли смо до закључка да домен вредности за атрибут спеловања могу бити ниске које садрже слова грчког алфабета и одабране дијакритичке симболе (акценте,



Слика 16. Главни формулар леме ΛΙΩΣΤΡΑ са акцентованим IPA симболима у пољу *Phonetic Type*

цртице, заграде и апострофе). Овакав закључак важио је за прву имплементирану верзију система. Од тренутка када је систем ушао у фазу израде, почели смо да примамо веома специфичне захтеве за побољшањима која су се тицала атрибута спеловања. Корисници су указивали да неке речи, будући да потичу из других језика, имају самогласнике који не постоје у стандардном систему спеловања циљног језика, а то је савремени грчки. Будући да дијалекти који су били у фокусу истраживања имају велики број позајмљеница из других језика, ове захтеве корисника нисмо могли да занемаримо. Зато је било неопходно да смислимо како да омогућимо корисницима система да разумеју како се изговарају дијалекатске речи, а да не морају да ишчитавају фонетске IPA симболе. Најједноставнији начин да се то постигне био је да се дозволи уметање графемских представа (слова) које се користе у језику порекла позајмљеница да би се представили они самогласници који не постоје у циљном језику (савременом грчком). Резултат наведеног био је мањи број самогласника са графемским представљањем са прегласом (u, e и

α са прегласом – ü, ë и ¨α). Ова слова била су укључена у једну виртуалну тастатуру предвиђену за уређивање атрибута спеловања (фонетске транскрипције). Слика 17 представља ову виртуелну тастатуру.

Слика 17. Виртуелна тастатура за графеме које представљају додатне самогласнике

4. Као што је описано у ранијем излагању, свака лема може имати различите реализације и сваку од њих карактерише унеколико различита фонетска реализација која зависи од микродијалекатског подручја са којег потиче. Овим смо се руководили приликом осмишљавања и изградње система у којем сваку реализацију карактерише једна микродијалекатска зона која представља уже подручје у оквиру шире дијалекатске регије. Међутим, у фази изградње система корисници су указивали да поједине реализације могу да постоје у више микродијалекатских зона (у оквиру главног дијалекатског подручја). Како би изашли у сусрет оваквим касно изреченим захтевима модификовали смо нацрт шеме података и саму апликацију. Атрибут *Microdialectal region* (микродијалекатска зона) модификован је тако да може да прими више вредности а одговарајућа GUI ставка је измењена како би обухватила листу вредности, при чему домен за сваку појединачну вредност чини скуп микродијалекатских зона које постоје за дијалекат одређене леме. Слика 18 приказује реализације понтске леме ΟΜΜΑ-ΤΟΤΖΑΤΖΙ за коју трећа наведена реализација (трећи ред) има три микродијалекатске зоне: Τραπεζούντα, Χαλδία, Σάντα.

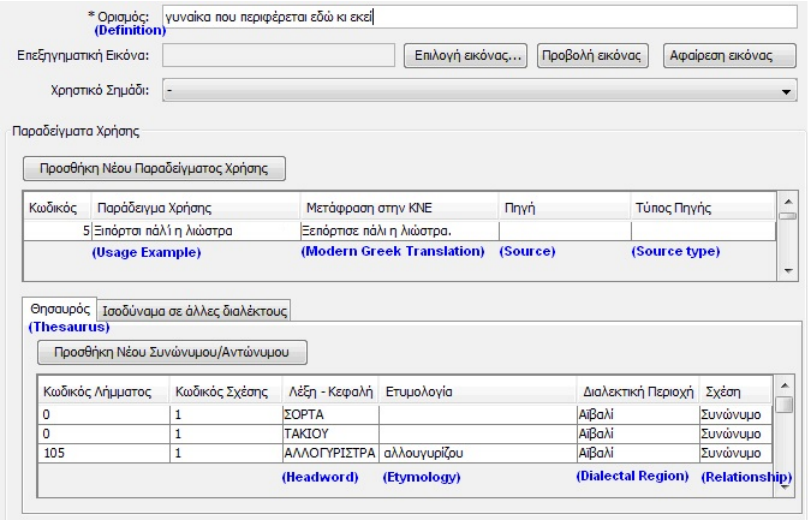
Τύποι Πραγμάτωσης | Σημασίες
(Realizations)

Δημιουργία Νέου Τύπου Πραγμάτωσης

Κωδικός	Φωνητικός Τύπος	... Φωνητική Ορθογραφία	Μικροδιαλεκτικές Περιοχές
4396	omatodʒádʒ	ομματοτζάτζ	Χαλδία
4397	matodʒádʒi(n)	ματοτζάτζι(ν)	Οινόη
4398	matodʒádʒ	ματοτζάτζ	Τραπεζούντα, Χαλδία, Σάντα
4399	matodʒǽci	ματοτζάϊ	Αμισός
4400	matodʒídʒ	ματοτζίτζ	Κοτύωρα
	(Phonetic type)	(Spelling)	(Microdialectal regions)

Слика 18. Реализације понтске леме ΟΜΜΑΤΟΤΖΑΤΖΙ

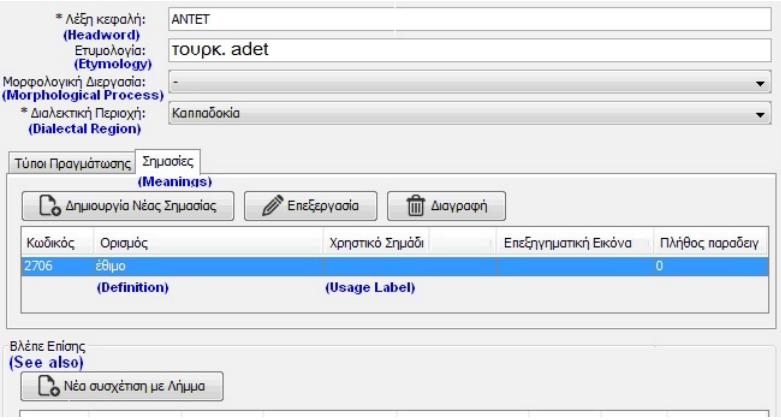
5. Још једно својство система које није било дефинисано у фази анализе али се појавило током развојне фазе било је садржај листе референци *see also* (види и) а која се односи на циљну леме која може да има више значења или више типова реализације. Решење за које смо се определили било је да представимо дефиницију сваког значења реферисане леме и спеловање сваке реализације реферисане леме у листи ставки „види и“. Доњи део слике 16 представља празну листу референци леме али може се видети да је множина коришћена у насловима релеватних колона – *Spellings* (Спеловања) и Meanings (Значења).
6. Још једно својство представљеног система вредно помена је да везе за релације синонимије и антонимије могу да упућују на леме истог дијалекта које могу а и не морају бити присутне у систему. Слика 19 представља значење леме ΛΙΩΣΤΡΑ са припадајућим „тезаурусом“, то јест табелом која садржи везе синонимије и антонимије. На овој слици можемо видети три везе које се односе на леме ΣΟΡΤΑ, ΤΑΚΙΟΥ и ΑΛΛΟΥΓΥΡΙΣΤΡΑ од којих је само трећа заступљена у систему. Поређењем слика 10 и 19 може се увидети да смо у табели тезауруса новије верзије (доњи део слике 19) заменили *Spelling* (Спеловање) са *Etymology* (Етимологија). Треба напоменути да је такође додата и нова колона – *Source type* (Тип извора) у табели примера употребе (средишњи део слике 19).



Слика 19. Формулар значења леме ΛΙΩΣΤΡΑ – значење и припадајући тезаурус

7. До сада нисмо дали ниједан пример еквивалента у другом дијалекту. Слика 20 је главни формулар кападочанске леме ΑΝΤΕΤ која има једно значење. На слици 21 дат је доњи део формулара значења који приказује картицу *equivalent in other dialects* (еквивалент у другим дијалектима). Као што се може видети на слици 21 једино доступно значење кападочанске леме ΑΝΤΕΤ има еквивалентно значење у ајваличкој леми ΑΝΤΕΤΙ.

8. Током фазе изградње описаног система приметили смо да су корисници одступали од прописа за писање примера употребе лема. Корисници су обично користили *copy/paste* могућност у оквиру оперативног система како би унели карактере које апликација није директно обезбеђивала за атрибут *usage example* (пример употребе). На пример, вредност ς̓λαγεύω το μήλον унета је као пример употребе за лему ΑΣΛΑΕΥΩ. Ова вредност има грчке карактере који одговарају правилима али такође садржи и један турски карактер (други карактер у ниски). Фраза Aς̓λαγεύω το μήλον

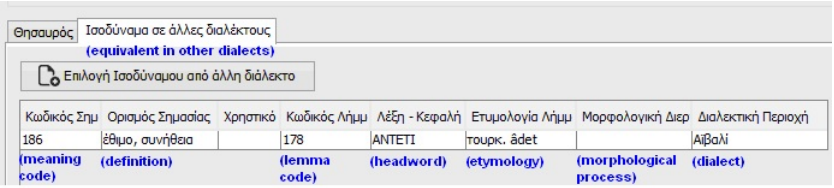


Слика 20. Главни формулар кападочанске леме ANTET

као вредност за пример употребе, заједно са вредношћу “τουρκ. Aşlamak” (то јест, „од турског aşlamak“) у оквиру атрибута *etymology* (етимологија) исте леме може да указује на то како би матерњи говорници дијалекта могли да напишу дијалекатску реч у њиховим документима (Açλαγεύω). Међутим, ова индикација скривена је унутар једног од значења леме. Претпостављамо да би било боље уколико би се обезбедио други атрибут, назван *indicative writing* (индикативни испис), за сваку реализацију леме. На овај начин поменути „индикативни испис“ био би директно доступан у главном формулару леме, а уз то би се могао издвојити за сваку појединачну реализацију леме (микродијалекатски регион). Ово је једино својство које није имплементирано у завршној фази оптимизације система јер је потреба за њим уочена сувише касно, али сматрамо да би оно било од значаја за наредне вишедијалекатске речнике.

8. Унапређење и проширење система

Шема података (ERD) за подршку унапређеног система приказана је на слици 22.



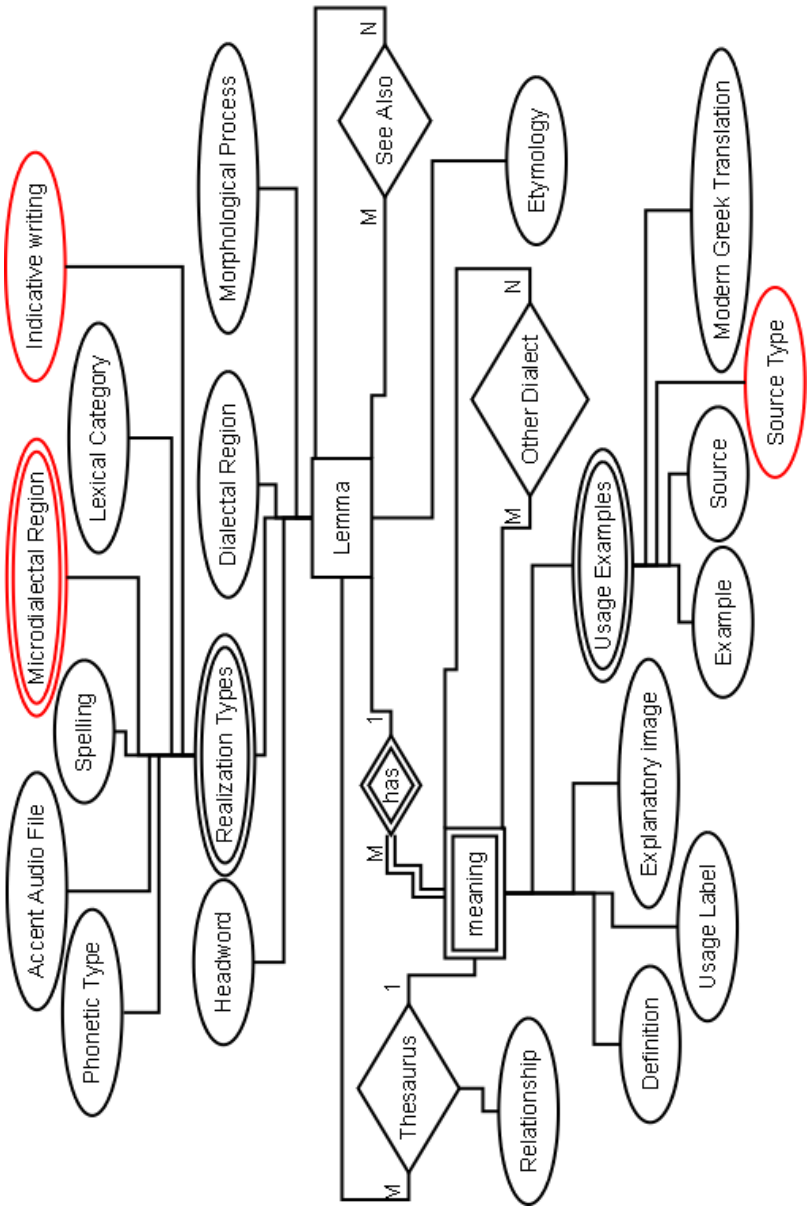
Слика 21. формулар значења леме ANTET – еквиваленти у другим дијалектима приказани су у предњем плану

9. Закључак

Планирана макроструктура система TDGDAM обухвата око 2.500 одредница за сваки од три дијалекта малоазијског грчког (што је укупно око 7.500 одредница). Ове одреднице прикупљене су искључиво из три поменута дијалекта и оне не обухватају елементе речника савременог грчког језика, осим у појединачним случајевима одступања од стандардне употребе. Њихово навођење дато је у складу са алфабетским а не ономасиолошким принципом, док је доступност обезбеђена преко динамичких опција претраживања (Xydopoulos, 2012).

Захвалност

Истраживање је финансијски подржано од стране Европске Уније (Европски социјални фонд, енг. European Social Fund – ESF) и грчких националних фондова кроз програм „Образовање и целоживотно учење“ који спроводи Национални стратешки референтни оквир – Програм за подршку истраживању: „Талис – улагање у друштво знања“. Изражавамо захвалност Ангели Рали, професору лингвистике на Одсеку за филологију Универзитета Патрас, која је координатор целокупног AMiGre пројекта. Дубоку захвалност дугујемо Георгију Ксидопулосу, ванредном професору лингвистике на Одсеку за филологију на Универзитету Патрас, који је поставио лингвистичке захтеве за дијалекатски електронски речник.



Слика 22. Унапређена шема података (ERD)

Литература

"The Dictionaries of ILSP". Institute for Language and Speech Processing. Приступљено 22.12.2016. http://www.xanthi.ilsp.gr/dictionaries/

"Homonym". Wikipedia. Приступљено 22.12.2016. https://en.wikipedia.org/wiki/Homonym

"Online Dictionaries". Portal for the Greek Language. Приступљено 22.12.2016. http://www.greek-language.gr/greekLang/modern_greek/tools/lexica/index.html

"Opposite (semantics)". Wikipedia. Приступљено 22.12.2016. https://en.wikipedia.org/wiki/Opposite_(semantics)

"Polysemy". Wikipedia. Приступљено 22.12.2016. https://en.wikipedia.org/wiki/Polysemy

"Synonym". Wikipedia. Приступљено 22.12.2016. https://en.wikipedia.org/wiki/Synonym

Barbato, Marcello и Alberto Varvaro. "Dialect dictionaries". *International Journal of Lexicography* Vol. 17 no. 4 (2004): 429–439

Béjoint, Henri. *Modern lexicography: An introduction.* Oxford: Oxford University Press, 2000

Burke, Sean. M. "The Design of Online Lexicons". У *A practical guide to lexicography*, Piet van Sterkenburg, 240–249. Amsterdam: John Benjamins, 2003.

Durkin, Philip. "Assessing Non-standard Writing in Lexicography". У *Varieties of English in writing: The written word as linguistic evidence*, Raymond Hickey, 43–60. Amsterdam: John Benjamins, 2010.

Geeraerts, D. "Principles in Monolingual Lexicography". У *Wörterbücher / Dictionaries / Dictionnaires. Ein internationales Handbuch zur Lexikographie. An International Encyclopedia of Lexicography. Encyclopédie internationale de lexicographie*, Franz Josef Hausmann et al., 287–296. Berlin: Walter de Gruyter, 1989.

Karanikolas, Nikitas N., Eleni Galiotou, George J. Xydopoulos, Angela Ralli, Konstantinos Athanasakos et. al.. "Structuring a Multimedia Tri-dialectal Dictionary". У *LNAI 8082: International Conference on Text, Speech and Dialogue*, Ivan Habernal и Vaclav Matousek, 509–518. Berlin Heidelberg: Springer, 2013.

Keymeulen, J. Van. "Pilot project: A Dictionary of the Dutch Dialects". Paper presented at the 14th EURALEX International Congress, Fryske Akademy, Ljouwert/Leeuwarden, 6–10 July 2010

Landau, Sidney. *The art and craft of lexicography (2nd edition).* Cambridge: Cambridge University Press, 2001

Markus, Manfred и Reinhard Heuberger. “The architecture of Joseph Wright’s English Dialect Dictionary: preparing the computerized version”. *International Journal of Lexicography* Vol. 20 no. 4 (2007): 355–368

Papadopoulos, A. *Historical Dictionary of the Pontic Dialect.* Athens: Epitropi Pontiakon Meleton, 1958

Penhallurik, Rob. “Dialect Dictionaries”. У *The Oxford History of English Lexicography Volume II*, A. P. Cowie, 290–313. Oxford: Oxford University Press, 2009.

Rys, Kathy и Jacques Van Keymeulen. “Intersystemic correspondence rules and headwords in Dutch dialect lexicography”. *International Journal of Lexicography* Vol. 22 no. 2 (2009): 129–150.

Themistocleous, C. “Cypriot Greek Lexicography: An Online Lexical Database”. Paper presented at the 15th EURALEX International Congress, University of Oslo, Oslo, 7-11 August 2012.

Thomason, Sarah G. *Language Contact. An Introduction.* Edinburgh: Edinburgh University Press, 2001.

Thomason, Sarah G. и Terrence Kaufman. *Language Contact, Creolization, and Genetic Linguistic.* Berkeley: University of California Press, 1988.

Tsalidis, Christos, Mavina Pantazara, Panagiotis Minos и Elena Mantzari. “NLP Tools for Lexicographic Applications in Modern Greek”. У *eLexicography in the 21st Century: New Challenges, New Applications (Proceedings of eLex2009)*, S. Granger и M. Paquot, 457–462. Louvain-la-Neuve: UCL Presses, 2010.

Xydopoulos, George J. “Metalexikografikes paratiriseis sta leksika Benardi ke Syrkou (Metalexicographic comments to Benardi and Syrkou dialectal dictionaries)”. *Patras Working Papers in Linguistics* Vol. 2 no. 1 (2012): 96–113

Xydopoulos, George. J. и Angela Ralli. “Greek dialects in Asia Minor: Setting lexicographic principles for a tridialectal dictionary”. Paper presented at the 5th MGDLT Conference, Ghent, Belgium, September 2012

Zgusta, Ladislav. *Manual of lexicography.* The Hague: Mouton, 1971

Заштита и промоција осетљивог архивског материјала из приватне Библиотеке Лазић[1]

УДК 930.25:004.9

Василије Милновић
milnovic@unilib.rs

Милена Костић
mkostic@unilib.rs

Матеа Милошевић
milosevic@unilib.rs

Универзитетска библиотека „Светозар Марковић“

САЖЕТАК: Предмет овог рада јесте упознавање са могућностима нових дигиталних технологија у заштити и презентацији осетљивог архивског и другог библиотечког материјала у контексту концепта „паметне библиотеке“ (*smart libraries*), као и, с тим у вези, прилог предлогу нацрта за стандардизацију процеса дигитализације те грађе. Савремене дигиталне технологије и дигиталне алатке дају нове могућности истраживачима из поља хуманистике који сада на модеран и транспарентан начин могу истраживати ретке и осетљиве архивске материјале. Овај рад ће показати да архивски материјал, и у ери „Индустрије 4.0“ (*Industry 4.0*) и „Интернета ствари“ (*Internet of Things*) може бити изузетно важан део промоције једне културе у контексту паметних библиотека. Ипак, посебна пажња биће посвећена стандардизацији дигитализације која се тренутно спроводи у Универзитетској библиотеци.

КЉУЧНЕ РЕЧИ: паметне библиотеке, нове дигиталне технологије, архивски материјал, библиотечки материјал, дигитализација, стандардизација.

РАД ПРИМЉЕН: 15. јул 2016.
РАД ПРИХВАЋЕН: 4. новембар 2016.

[1] Рад настао у оквиру пројекта Safeguarding the fragile collection of the private archive of the Lazic family, за који је Универзитетска библиотека “Светозар Марковић” добила грант Британске библиотеке из програма Endangered Archives

1. Увод

Концепт паметних библиотека се интензивно развија у ери „индустрије 4.0“ где „интернет ствари“ постаје свепрожимајући (Evans, 2011). Улога паметних библиотека у том контексту јесте да постану главна информативна средишта у оквиру којих се експериментише са новим технологијама и упознаје са концептима везаним за технологију. Данас су библиотеке центри учења нових технологија и истраживања помоћу истих, а све ради продуктивније, концизније, свеобухватније и лакше истраживачке делатности. Пре свега, библиотеке представљају улаз у свет дигиталног. Оне су предвиђене каоместа у којима се најпреимплементирају нове технологије (Min, 2012; Younis, 2012), које касније могу имати ширу привредну или јавну примену, па су тако библиотеке и потенцијални економски инкубатори. Представљају најлепши начин да се пригрле нове технологије и да се избегне дигитални егзил.

Основни циљ пројекта *Safeguarding the fragile collection of the private archive of the Lazic family* јесте трајно похрањивање и заштита изузетно вредне колекције која се налази у приватном власништву Удружења грађана „Адлигат“ (Библиотека Лазић). Колекција која ће бити дигитализована и приказана јавности састоји се од неколико потколекција: правне публикације, ратне публикације, периодика, календари и архивска грађа. Овај материјал има и додатну историјску вредност у склопу актуелног обележавања стогодишњице Првог светског рата. Поред тога, овај материјал је изузетно релевантан за високошколске библиотеке будући да је једним делом састављен од униката и ретких издања која се не могу наћиу фондовима других библиотека. Шира и стручна јавност ћепрви пут бити у прилици да види *Преглед листова*, тајни часопис Владе Србије у егзилу штампан у Женеви, као и преглед актуелних вести са савезничке и непријатељске стране. Материјал је драгоцен широком кругу истраживача: од историчара и социолога, преко антрополога и филолога, до библиотекара и архивара.

Посебно је важно нагласити да је у питању заједничко прегнуће јавног и невладиног сектора, то јест,академске институције – Универзитетске библиотеке „Светозар Марковић“ и удружења грађана „Музеј књиге и путовања – Адлигат“. Овај вид сарадње значајан јекао савремена форма пројектних активности која ће бити заокружена учешћем приватног сектора. Изашавшиизстрогих оквира своје примарне

делатности, библиотека иде у сусрет потребама других организација и струка, што је – укључујући нужни мултидисциплинарни приступ – важноу контексту даљег напретка друштва, концепта „паметних градова“ (*smart cities*) и паметних библиотека, као централних информативних пунктова (*hubs*) будућег друштва знања.

2. Пролегомена Универзитетске библиотеке ка концепту „паметних библиотека“

Уз пуну свест о значају отвореног приступа за достизање „друштвазнања“ (*knowledge society*) и улози високошколских библиотека у дисеминацији знања, ови изузетно осетљиви материјали биће представљени корисницима посредством нове дигиталне технологије Magic Box, која је погодна за презентацију и интерактивни приказ осетљивогматеријала. Претраживање дигиталних садржаја је транспарентно, преко екрана осетљивог на додир, док се конкретна публикација истовремено може посматрати иза екрана у аутентичном стању. Поред дигитализације штампаних публикација, овај уређај омогућава корисницима да прегледају галерије слика, 3Д приказе и филмове. Овај екран обезбеђује јединствено искуство заинтересованима за ретке или осетљиве публикације чији је приступ обично ограничен. Универзитетска библиотека „Светозар Марковић“ је прва институција у источној Европи и трећа библиотека на свету која поседује овај уређај, који ће постати својеврсни пулсирајући прозор у свет интересовања стручњака из различитих хуманистичких дисциплина.

Према уговору потписаном између Универзитетске библиотеке „Светозар Марковић“ и Британске библиотеке, сви дигитализовани материјали биће доступни на сајту Британске библиотеке[2], као и у оквиру посебног дигиталног репозиторијума Универзитетске библиотеке, чија је израда у току. С обзиром на план пројекта, мастер фајлови чуваће се у tiff формату, одвојено на *cloud* платформи – *Therefore*, у власништву Универзитетске библиотеке, док је у току припрема дигиталног репозиторијума са претраживим садржајем, и то тако што ће у овом будућем репозиторијуму корисницима бити доступне само слике докумената без контрастне позадине и калибрацијских картица. За

[2] http://eap.bl.uk/database/overview_project.a4d?projID=EAP833;r=9741#project_outcome

Слика 1. Magic Box

.

припрему ових колекција користиће сесофтвер *docWorks*, који је основни модел програмаза приређивање садржаја у *Magic Box*-у.

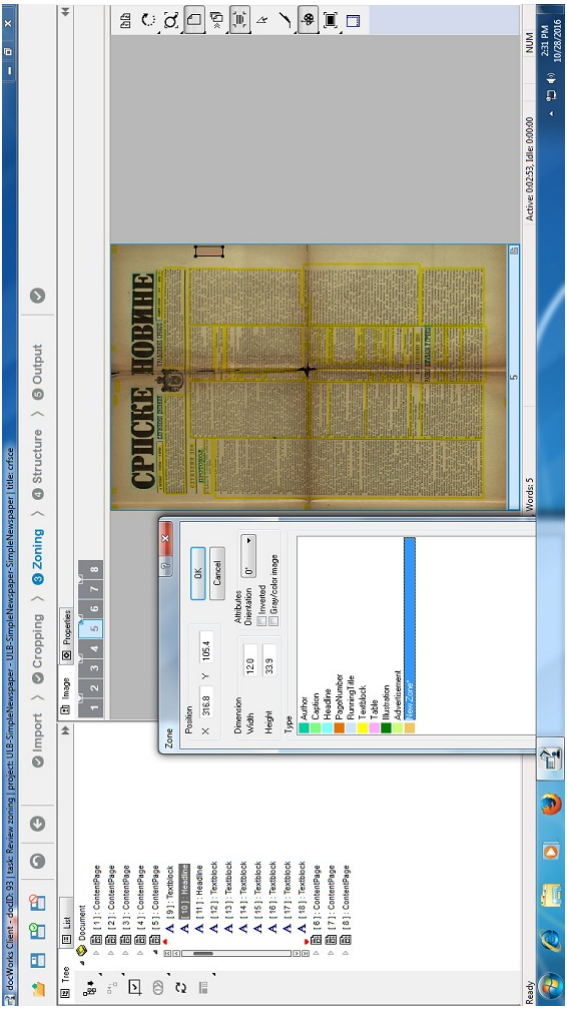
Припремање материјала у програму *docWorks*[3] обухвата следеће кораке:

1. исецање (*crop*) површина појединих страница дигиталних објеката;
2. зонирање објеката сегментирањем страница у блокове и ступце са дефинисаним површинама за OCR (*Optical Character Recognition*) и одређивањем њиховог типа у зависности од функције у објекту: наслови, текст, аутор, слике итд.;
3. сређивање структуре објекта (ставка, поглавље, чланак) повезивањем одговарајућих наслова и припадајућих садржаја;
4. исправљање текста и метаподатака;
5. стварање готових објеката у форми METS/ALTO фајлова прикладних за приказ у *Magic Box*-у и репозиторијуму.

Сви поменути кораци подразумевају прво аутоматску анализу, а тек потом ручно исправљање.

Будући да PDF нуди лимитиране могућности претраге по кључним речима, Универзитетска библиотека је усвојила METS/ALTO фајлове. Треба нагласити да смо прве овакве фајлове добили преко учешћа у пројекту *Europeana Newspapers*, а да смо потом успели први у Србији, уз наше колеге из Народне библиотеке

[3] http://content-conversion.com/#docworks-2



Слика 2. docWorks

.

Србије, да их самостално производимо у неограниченим количинама. Најједноставнија дефиниција ових фајлова јесте да преко њих институције своје дигиталне објекте могу прилагодити тако да добију претраживе документе. То је управо оно што ће се применити на дигитализоване материјале у оквиру поменутог пројекта.

METS и ALTO стандарди су установљени ради лакшег описа дигитализације штампаних материјала. Идеја је била да се раздвоје описне информације од самих садржаја материјала да би се дигиталним објектима руковало на једноставнији начин, јер када су сви подаци у једном XML фајлу (као што је случај код TEI – *Text Encoding Initiative* формата), XML фајл је огроман.

METS (*Metadata Encoding and Transmission Standard*) је слободан стандард заснованна XML-у (XML based open standard) који је 2001. године установила Конгресна библиотека у Вашингтону и користи се за дугорочно чување фајлова којим се описују дигитални објекти, штампани медији (књиге, новине, часописи), аудио или видео материјали итд. METS обично садржи више врста стандарда метаподатака: описне, административне, структурне информације, затим стандарде физичке и логичке структуре, као и линкове ка другим дигиталним објектима, сликама, аудио/видео и текстуалним фајловима.

ALTO (*Analyzed Layout and Text Object*) је слободан стандард заснованна XML-у, који је такође основала Конгресна библиотека. Сврха му је дигитални опис распореда штампане стране тако да је могуће поново изградити оригиналну страницу. Овај фајл обухвата садржаје појединачне стране дигиталног документа и може садржати тагове који доносе више података о самом објекту. Описује стилове, распоред и тип блокова информација.

Дигитални објекти структурирани на овај начин биће много оперативнији за употребу и омогућиће јединствену претрагу – и у физичком простору, кад је реч о новој технологији *Magic Box* и онлајн, када је реч о специјализованом дигиталном репозиторијуму – са резултатима који ће кориснику омогућити детаљан преглед садржаја колекције, као и брзу и лаку претрагу кроз садржину преко кључне речи. Овоме треба додати да ће, осим самог садржаја дигитализовног објекта, кориснику бити доступни и метаподаци објекта – који су већ начињени – али и додатна експертска литература, када је реч о самој теми датог објекта. На тај начин књига није само дигитализована; такође је и датификована (*datafied*). Књиге постају data sets, односно text cor-

pora, а речи постају *data points*. Машине, дакле, постају читаоци (*readers*).

3. Прилог Универзитетске библиотеке предлогу нацрта дигиталног репозиторијума Републике Србије

Будући да је имплементација овог пројекта још у току, намера је да закључке пројекта представимона завршној стручној конференцији, уз учешће познатихстручњака и потенцијалних заинтересованих партнера из различитих области ка којима је пројекат усмерен. Тиме ће се широј јавности пружити могућност да активно учествује у оваквим активностима. Пројекат ће имати додатну вредност кроз заснивање шире, јавне дискусије о концепту паметних библиотека и „дигиталног грађанства“ (*digital citizenship*). Истовремено међународни пројекти су начин, не само да се у ограниченим условима финансирања јавних институција, нарочито оних из домена културе и науке, допуни буџет, већ и да се врши стручно усавршавање и напредовање запослених.

До сада се у Србији појам дигитализације најчешће поистовећивао са појмом скенирања материјала. Разуме се, по правилу се узимају у обзири међународни стандарди који су увек основа за израду општих препорука док се не произведе обједињени домаћи стандард, као што су *Technical Guidelines for Digitizing Cultural Heritage Materials*[4] или стандарди које препоручује Унеско[5]. У координацији са Британском библиотеком, Универзитетска библиотека „Светозар Марковић“ усваја нове концепте дигитализације који садрже и нове стандарде.

Поступци којима се до данас водила дигитализација књижне грађе у Србији створили су бројне дигиталне репозиторијуме[6] и доказали да је

[4] http://www.digitizationguidelines.gov/guidelines/FADGI_Still_Image_Tech_Guidelines_2016.pdf (приступљено 27. 10. 2016).

[5] http://www.unesco.org/new/fileadmin/MULTIMEDIA/HQ/CI/CI/pdf/mow/digitization_guidelines_for_web.pdf (приступљено 27. 10. 2016)

[6] Ово су неки примери домаћих дигиталних репозиторијума: Дигитална Народна библиотека Србије (http://www.digitalna.nb.rs/), Претраживе дигитализоване историјске новине (http://www.unilib.rs/istorijske-novine/pretraga), Дигитална библиотека Матице српске (http://digital.bms.rs/ebiblioteka/), Дигитална Библиотека града Београда (http://dlibra.bgb.rs/dlibra)

Србија спремна да успешно одговори на још један захтев информатичке револуције. Упркос силовитом темпу развоја и убрзаном напретку у овој области, кадрови оријентисани на рад и усавршавање у сфери дигитализације књижне грађе успешно одговарају ритмукоји диктирају најнапреднији међународни центри у овом домену. Ради успешног обављања послова дефинисани су и први нацрти стандарда који добијене дигиталне објекте треба да обједине у квалитету. Следећи дугогодишњу успешну праксу у дигитализацији књижне грађе, Универзитетска библиотека покушава да одговори на стандарде за дигитализацију Британске библиотеке, једног од светских лидера у овој области.

Сврха Програма за угрожену архивску грађу (Endangered Archives Programme – EAP) јесте заштита грађе кроз проналажење прикладних услова за њено чување и истовремену дигитализацију, која би обезбедила дугорочно складиштење и шири приступ дигиталним објектима (Endangered, 2016). Да би се одговорило на захтеве програма, изузетно је важно да дигиталне копије одговарају архивским стандардима. Неки стандарди које испред својих партнера поставља Британска библиотека већ се налазе у спискустандарда и успешно примењиваних пракси институција које се баве дигитализацијом у Србији, али неки стандарди постављају и нове изазове и помажу да у овој области– као чувари овдашње баштине и њени промотори у складу са најновијим концептима као што су Индустрија 4.0 и паметне библиотеке – будемо још успешнији и бољи.

Док су захтеви креирања дигиталних копија физичких објеката везани за резолуцију – минимум 300dpi (*dots per inch*), или формат tiff (*Tagged Image File Format*) – већ увелико примењивани и добро познати у овдашњој пракси, одговарајући на захтеве Програма за угрожену архивску грађу, упознали смо и нове праксе и стандарде те смо, на тај начин, спонтано радили на усавршавању сопствених смерница.

Пре свега, рад на овом пројекту, као прилог нацрту стандарда за дигитализицију старе и ретке књижне грађе у Универзитетској библиотеци, унео је две нове димензије: боју и дужину.

Као што географску карту (пример дводимензионалне слике географског подручја) ближе објашњавамо помоћу размере и боје, скен или фотографију (слику дигитализованог објекта) прецизније ћемо одредити лењиром (*ruler*), као размерском величином и калибрацијском картом боја (*color calibration card*), као додатним описом. На овај начин сваки скен или свака слика пружићепрецизније податке о стварном физичком изгледу објекта, дајући недвосмислени податак о прецизним

димензијама, а захваљујући калибрацијској карти боја или колор шеми, још прецизнији увид у колорит објекта.

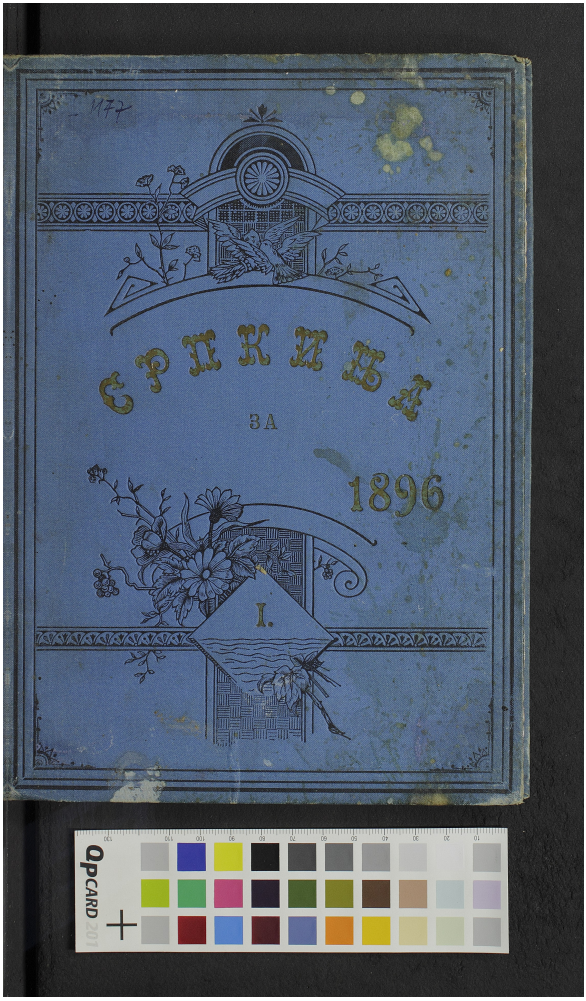
Разни извори светла имају различите температуре, па слике снимане у другачијим условима немају прецизне боје објекта. Да би се ово донекле избегло, користи се тзв. *White Balance*, или „баланс белог“, то јест корекција боја на основу извора. Овако кориговане боје мењају равнотежу између RGB (red, green and blue) кривих (RGB *curves*), али не и њихов облик и позицију, па се на слици битније мења осветљеност, али не и саме нијансе. Такође, због различитог начина репродукције или стварања слика на различитим уређајима не добијамо идентичне боје. Управо зато да би овако настале конверзије биле мање и слика квалитетнија, са што мањим разликама, развијено је управљање бојама (*Color Management*).

Једна од главних алатки за управљање бојама је калибрацијска картица боја или колор шема. Подешавање боје да одговара стварним нијансама на фотографији је један од изазова и у дигиталној и уаналогној фотографији. Шведска компанија *QP Cards AB* из Гетеборгаза ове потребе створила је исплативо и ефикасно решење које функционише на основубесплатног корективног софтвера и калибрацијских картица које се плаћају и доступне су у неколико верзија.

QP картица (*QPcard*) је само један од прихваћених модела калибрацијских картица. Ове картице се набављају као индустријски, тј. фабрички произведене, најчешће су од картонског материјала и није дозвољено њихово самоиницијативно штампање, посебно ако се користе да би се њима одговорило на неки стандард. Са једне стране обично имају лењир, док се целом дужином простире колор шема.

Корекција боја на сликама са QP картицом могућа је помоћу софтвера за калибрацију, *QPcolorssoft* 501, који се може преузети са сајта произвођача. Тим софтвером и QP картицом постављеном у сцену снимања – најбоље паралелно у односу на сензор фото-апарата или камере, тако да није постављена под углом или у некој сенци током снимања, креира се референтни профил. Пошто се фискира „баланс белог“ за дате услове и потом креира одговарајући профил боја са задатим вредностима, прави се референтни профил за корекцију и све наредне слике снимљене са QP картицом могуће је подвргнути калибрацији.

Калибрација се врши тако што се алатком Select обухвати (уоквири) QP картица, а затим фино подеси по свим бојама са картице, тако да



Слика 3. QPcard

.

свака заузме право место у формираном шаблону, а затим се креира одређени профил боја и референтни профил за калибрацију. Када се креира профил довољно је све следеће снимке, настале у истим условима осветљења и са истим балансом беле боје, подвргнути групној корекцији боја.

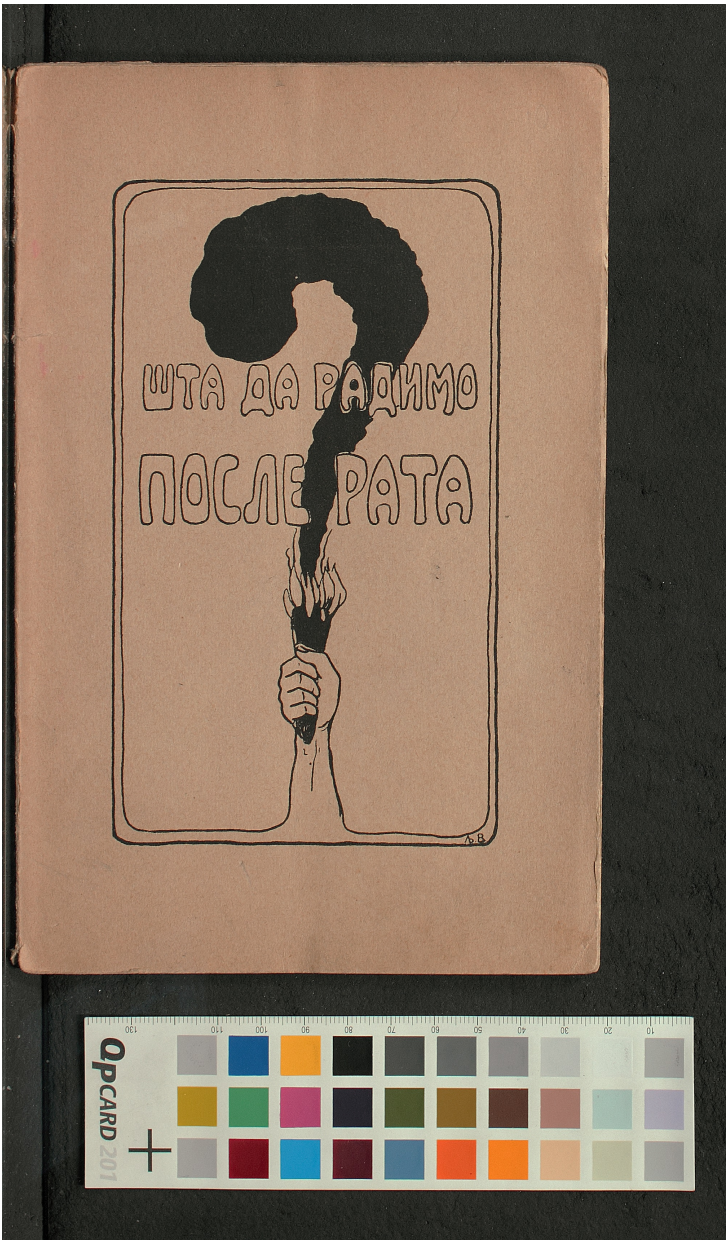
Када би се равноправно – поред tiff-а, као прикладног формата слика, и 300dpi, као основне минималне резолуције, довољне и неопходне за OCR, потпуно обухваћене површине сниманог објекта, такорећи од ивице до ивице – у сваку слику дигиталног објекта укључила и једна колор шема са лењиром, онда би овако креиран дигитални објекат био готово потпуно идеална слика реалног.

Разуме се, читав процес дигитализације, као и сам потенцијални будући стандард, мора поседовати спровођење контроле квалитета и поступак евалуације дигитализације. У оквиру нашег пројекта, Универзитетска библиотека се руководила упутством EAP програма и спроводила следеће мере:

1. на крају сваког радног дана неопходно је извршити контролу квалитета направљених скенова;
2. скенови се копирају на екстерни хард-диск и чувају на различитим локацијама (back up);
3. када се слике сачувају треба проверити да ли су све ротиране тако да је материјал читљив;
4. непосредно пре коначног трајног похрањивања материјала примењује се *MD5 checksum* за детекцију грешака.

MD5 checksum, што у дословном преводу значи „контролна сума“, представља 128-обитну суму, попут отиска прста за одређени фајл, са изузетно малом вероватноћом да се добију две идентичне вредности контролних сума за два различита фајла.[7] То је изузетно важноза поређење самих фајлова, као и за контролу њиховог интегритета. Да бисмо разумели како функционише ова вредност, треба претпоставити да постоје физички раздвојене две сличне огромне датотеке за које треба утврдити да ли су међусобно исте или различите, а да их притоме није могуће спојити и упоредити директно. Помоћу *MD5 checksum* – а довољно је израчунатипојединачне контролне суме за обе датотеке, а

[7] http://www.fastsum.com/support/md5-checksum-utility-faq/md5-checksum.php



Слика 4. Пример корекције помоћу QPcolorssoft 501

потом их упоредитида би се добио недвосмислен податак о томе да ли је реч о истом или различитом објекту.

Примена добрих пракси успешних светских центара за дигитализацију може бити чврст темељ за озбиљније дефинисање стандарда на нивоу институције, али и дигитализације књижне грађе једне државе у целини. На овај начин, стандардизовано снимање дигиталних објеката могло би успешно да послужи као основа за уједначавање квалитета дигитализованих материјала појединачних институција. Уједначавање дигиталних објеката и међусобна сарадња институција ујединили би донекле распршену енергију дигиталних центара у земљи, што би резултирало јединственим и обједињеним дигитални репозиторијум на националном нивоу.

4. Значај и садржина дигиталне колекције у оквиру Endangered Archives пројекта

У оквиру колекције Библиотеке Лазић, која је дигитализована уEndangered Archives пројекту Универзитетске библиотеке, налази се неколико већ поменутих потколекција. Дигитална збирка, обухваћена пројектом, броји 50.055 дигитализованих страница.

4.1 Потколекција „Правне књиге“

Књиге из области права које су изазвале велико интересовање комисије Британске библиотеке за додељивање Endangered Archives гранта, могу се, условно, поделити у две групе: оне које је објавио Геца Кон и остале правне публикације.

Геца Кон је најважнији српски и југословенски издавач на почетку XX века, хапшен током Првог светског рата од аустроугарске војске због глорификације победа српске војске,а током Другог светског рата,као припадник јеврејске заједнице у Србији, са остатком породице затворен је у логоргде је и умро. Издавачка делатност Геце Кона је у одређеним моментима била већа од свих других издавача у Србији заједно. Неки најбитнији наслови српске књижевности, историје или права објављени су управо захваљујући овом великану. Узимајући у обзир његову активност у Првом светском рату и јеврејско порекло, његове публикације биле су масовно уништаване током оба светска рата. Тако је, примера ради, аустроугарска војска током окупације „уприличила“

јавно спаљивање његових издања, док су га у Другом светском рату Немци узимализа пример антијеврејске пропаганде, нарочито током 1941. Књиге из његове књижаре конфисковане су и прослеђене у Беч. Током самог рата, издавачку компанију преотела му је фашистичка издавачка кућа „Југоисток“, да би, након ослобођења, ова компанија постала „Просвета“ – најважнија државна издавачка кућа између 1950. и 1990. Такође, након ослобођења, његове публикације сматране су непожељним, јер су доживљаване као дело „класног непријатеља“. Зато се нека његова издања данас сматрају ретким. То нарочито важи за поједине публикације из области права, које притом никада нису дигитализоване, а веома често су у јако лошем физичком стању.

Будући експертска литература, ове публикације нису штампане у великом броју примерака, а с обзиром на то да су њихови предмети најчешће одређени закони из доба предратног капитализма, оне су додатно уништаване, па данас спадају у ретка издања. Посебну вредност даје чињеница да готово све публикације из ове потколекције потичу из личних библиотека људи важних за српско право и државу. У том смислу, белешке и *Ex Libris*-и су често једини траг постојања појединих књига и представљају вредност за себе. Овде можемо поменути само неке наслове:

1. „Политичке и правне расправе“, Слободан Јовановић, 460 страница, 1910. Изузетно ретка публикација једног од најеминентнијих српских интелектуалаца, који је у доба егзила током Првог светског рата био председник Прес-бироа српске владе, а касније председник Српске краљевске академије и ректор Универзитета у Београду, професор јавног права и декан Правног факултета. Као припадник старог естаблишмента, његова литература била је уништавана након Другог светског рата. Ко год је у кући имао његове публикације, ризиковао је да буде испитиван у полицији. Рехабилитован је тек 2007. Овај примерак припада легату професора Петра Бингулца (1897–1990), у власништву Библиотеке Лазић.
2. „Уговор о продаји и куповини, предавања на Правном факултету“, Живојин М. Перић, 149 стр, 1920. Изузетно ретка публикација најпознатијег српског адвоката између два рата. Овај примерак има ознаку Библиотеке Врховног суда.
3. „Првобитно словенско право пре X века“, Карло Кадлец, 130 стр. Изузетно ретка студија.

Укупан број публикација у овој потколекцији је 132. Ниједна публикација никада пре није била дигитализована, па њихову дигитализацију жељно ишчекују страни експерти и шира јавност.

Други део ове потколекције чини 29 ретких правних издања која нису објављена код Геце Кона. То су неке од следећих публикација:

1. „Измене и допуне уредбе о ратним инвалидима и други прописи о инвалидима“, 1938, 130 стр. Изузетно ретка публикација, нарочито важна због ретроспективног погледа на положај и третман ратних ветерана.
2. „Закон о уређењу Министарства иностраних послова и дипломатских и конзуларних заступништава Краљевине Југославије у иностранству“, 32 стр, Државна издавачка кућа Краљевине Југославије, Београд, 1929. Изузетно ретка публикација за официјалну употребу коју су користила сва југословенска конзуларна представништва у Европи, где су, поредосталих, радили и неки најистакнутији представници српске књижевности и културе: Иво Андрић, Милош Црњански, Јован Дучић. Овај примерак припада легату професора Петра Бингулца.
3. „Збирка закона новог доба прокламација Њ. В. Краља од 6. јануара 1929. године“, 67 стр, коју је штампао др Часлав М. Никитовић 1929. Изузетно ретка публикација. Занимљиво је да су ови закони остали на снази и у време увођења диктатуре. Припада легату судије Слободана Ћирића.

4.2 Потколекција „Ратне публикације“

С обзиром на тадашње трагичне догађаје, српска ратна издања (1914–1918) представљајуспецифичан библиотечко-архивски материјал. Након победа на Церу и Колубари, Србија је нападнута са трију страна, па је, након очајничких борби, одлучено да се цеодржавни апарат, влада, парламент, војска и огроман део српског народа повуче у Грчку. Будући да се војска кроз албанске планине повлачила зими, мање од 200.000 људи завршило је ово трагично путовање. Тако се дошло до јединственог положаја у историји: да једна држава, са владом, парламентом и законодавством, настави да постоји, али без територије. Један од најважнијих доказа тог континуитета јесте ратна издавачка делатност, па не треба да чуди што се све публикације штампане на Крфу (где је било седиште српске владе), у Бизерти у Тунису (где

је боравио велики број рањеника) и у Солуну, према нашем закону, третирају као национално културно добро. Ово важи и за поједине ретке публикације објављиване међу српском емиграцијом у Женеви, Ници, Лондону и у САД. Библиотека Лазић има једну од најдрагоценијих збирки ратних публикација, а у оквиру Endangered Archives пројекта биће дигитализовано 156. Међу њима су:

1. „Енглеско-српски речник“, Ђорђе А. Петровић, 192 стр, Солун, 1918. Изузетно ретка публикација, коју су користили војници да би комуницирали са енглеским лекарима.
2. „Закон о изменама и допунама закона о Народној банци“, 5 стр, Крф, 1916. Изузетно ретка публикација коју не поседује ниједна друга библиотека у Србији.
3. „Тајна превратна организација извештај са претреса у војном суду за официре у Солуну по белешкама вођеним на самом претресу“, 638 стр, Солун, 1918. Изузетно ретка публикација о окривљеним официрима за покушај атентата на српског престолонаследника који су претходно смакли краљевску династију Обреновић 1903. године. Ово је највећи судски процес који се одиграотоком рата.
4. „Српски школски дан у Француској 13/26 марта 1915 године“, 242 стр, Ниш, 1915. Веома ретка публикација писана у време повлачења српске војске. Ова публикација требало је да послужи каодоказ да српска војска има подршку савезника.

4.3 Потколекција „Периодика“

1. *Преглед листова* је публикација која ће први пут бити јавно приказана. Реч је о тајном информативном часопису Владе Србије у егзилу, објављиваном у Женеви, у малом тиражу (вероватно, мање од 50 примерака, од којих за само трипостоје сазнања да су сачувана). Занимљиво је да се часопис штампао на различитим типовима папира, на хектографу, зависно од тренутне ратне ситуације. „Преглед листова“ је заправо сачињен од одабраних информативних и новинских текстова из савезничке и непријатељске штампе, о тренутним ратним активностима и медијској слици Србије. Означено као „строго поверљиво“, било је намењено само највишим члановима владе и највишем официрском кадру. Највећи број примерака је неповратно уништен. Четири и по хиљаде сачуваних страница, у поседу Библиотеке Лазић, овом

дигитализацијом ће, у складу са највишим стандардима, бити први пут представљено српској јавности.

2. *Мисао* је изузетно редак српски ратни часопис штампан у Енглеској. Постоје само четири броја овог часописа и сви су у поседу Библиотеке Лазић. Уређивали су их тамошњи српски интелектуалци. Сва четири броја су дигитализована.
3. *Крфске новине*, 500 стр. Веома важна публикација јер су у њој, у току рата, српски књижевници, нарочито песници, објављивали своје радове. Касније ће ово стваралаштво бити сматрано најзначајнијим српским ратним стваралаштвом. Овај часопис је у саставу легата Стевана Бешевића, једног од уредника.

4.4 Потколекција „Календари“

Ово је према речима комисије за доделу гранта Британске библиотеке веома драгоцена колекција. Календари су увек били омиљена периодика, објављивани једном годишње, најчешће на хартији слабог квалитета, у облику књиге сачињене од најразличитијих садржаја из популарне науке и забаве, замишљене као штиво за читаву годину. Најчешће су уништавани на крају године, па су због тога, као и због слабог квалитета папира, данас изузетно ретки. Дигитализација неких календара из ове потколекције, нарочито оних из XIX века, представља прави стручни изазов услед изузетне осетљивости материјала. Управо због тога, власници оваквих публикација не желе да их прикажу широј јавности у физичком облику, чак ни у прилагођеним условима, па ће јавност овим путем први пут бити у могућности да види неке од календара.

1. „Вардар: календар за 1898“ – календар за 1898, 62 стр, издање књижарнице Љ. Јоксимовића, 1897. Изузетно редак календар који није у фонду Народне библиотеке Србије.
2. „Орао: велики илустровани календар за годину 1889. која је проста, има 365 дана“ – илустровани календар за 1889, за 365 дана, 176 стр, Нови Сад. Веома ретка публикација, значајна јер се у њој налазетекстови познатих српских писаца.
3. „Забавни календар за 1855“, Светозар Стојадиновић, Земун, 67 стр, написан пре реформе српског писма 1868. Изузетно је редак и сматра се културним добром од изузетног националног значаја.
4. „Српкиња: илустровани календар за наш женски свет за просту годину 1897“ – илустровани календар за даме за 1897 (издат 1896),

126 стр, Кикинда. Свакако једна од најређих публикација из ове потколекције, писана искључиво за даме. Ову потколекцију чине укупно 32 публикације.

4.5 Потколекција „Архивски материјал“

Овај архивски материјал се односи на Први светски рат и укључује: неколико писама, бележака, као и разгледница из Првог светског рата, фотографију војника на дан капитулације Бугарске, неколико кратких писама, изузетно ретке разгледнице-фотографије Српског избегличког позоришта из Бизерте (Тунис) итд. Овај материјал посебно је значајан из угла актуелне стогодишњице Првог светског рата. Ова потколекција је посебно обогаћена изузетно вредном колекцијом писама и телеграма, упућених породици (пре свега, супрузи Лујзи) војводе Живојина Мишића, поводом његове смрти, 1921. године. Збирка обухвата и говор одржан на самој сахрани. Већина поменутог архивског материјала никада није била ни објављена, а камоли дигитализована, па ће ово бити прилика да се стручна и шира јавност ближе упозна са овим материјалом. Укупан број јединица у потколекцији је 105.

5. Закључак

Из свега реченог, јасно је да се промоција и презентација националне баштине и њихова актуелизација, кроз савремене технологије и најновије концепте паметних библиотека, „интернета ствари“ и „индустрије 4.0“, не може адекватно спровести без ширег оквира и испуњавања стандарда на националном нивоу. Прилог предлогу нацрта Универзитетске библиотеке за стандардизацију процеса дигитализације само је први корак ка извесном „оживљавању“ вредне архивске и библиотечке грађе на нашим просторима. На актуелан и технолошки супериоран начин указује се на значај баштине, а како је овај процес део међународног пројекта, својеврсна вредна домаћа архивско-библиотечка грађа представљена је на недвосмислен, компаративан и транспарентан начин у оквирима заједничког европског наслеђа.

Литература

"Endangered Archives Programme. Guidelines for photographing and scanning archival material". http://eap.bl.uk/downloads/guidelines_copying.pdf. Приступљено 1.7.2016.

Evans, Dave. "The Internet of Things: How the Next Evolution of the Internet Is Changing Everything", 2011. http://www.cisco.com/c/dam/en_us/about/ac79/docs/innov/IoT_IBSG_0411FINAL.pdf. Приступљено 11.7.2016.

Min, Byung-Won. "Next generation library information service – smart library". *International Journal of Software Engineering and Its Applications* Vol. 6 no. 4 (2012): 171–194.

Younis, Mohammed I. "SLMS: a smart library management system based on an RFID technology". *International Journal of Reasoning-based Intelligent Systems* Vol. 4 no. 4 (2012): 186–191.

Отворени образовни ресурси и земље у развоју: један критички осврт

УДК 37.014(1-773)

Марија Пантић
p.mmmmmm@gmail.com
Универзитет у Београду
Филолошки факултет

САЖЕТАК: Овај рад се бави глобалним друштвено-економским контекстом појаве отворених образовних ресурса и њихове посебне улоге у образовним системима земаља у развоју. Централно место дато је самом појму образовања, који ће бити разрађен кроз одабране критике и анализе реформи образовних система, као и питања основних услова, метода подучавања и очекиваних резултата у образовању. Посебна пажња посвећена је глобалном сиромаштву, односно неједнаком глобалном развоју, а у складу с тим и све агресивнијој репродукцији постојећих друштвено-економских односа кроз неједнако образовање. Разматрање ових општих услова за резултат има неизбежно скептичан и критички приступ идеолошким, технолошким и педагошким аспектима иницијатива које креирају и промовишу отворене образовне ресурсе.

КЉУЧНЕ РЕЧИ: отворени образовни ресурси, земље у развоју, глобализација, неједнакост

РАД ПРИМЉЕН: 16. октобар 2016.
РАД ПРИХВАЋЕН: 23. децембар 2016.

1. Шта су отворени образовни ресурси?

Отворени образовни ресурси (Open Educational Resources (OER)) су материјали за подучавање, учење и истраживање који се налазе у јавном

домену или су издати са отвореном лиценцом која допушта бесплатан приступ, коришћење, прераду и дељење од стране других лица, без задржавања права или уз одређена права задржана[1].

Масачусетски институ за технологију 2001. године објавио је намеру да омогући бесплатан приступ свим својим универзитетским курсевима за некомерцијално коришћење, што је за крајњи исход имало појаву отворених курсева. Организација Уједињених нација за образовање, науку и културу (United Nations Educational, Scientific and Cultural Organization (UNESCO)) 2002. године сазвала је форум на тему утицаја отворених курсева на више образовање земаља у развоју. На овом форуму први пут је усвојен термин "отворени образовни ресурси".

Центар за истраживање и иновације у образовању (Centre for Educational Research and Innovation (CERI)) Организације за економску сарадњу и развој (Organisation for Economic Co-operation and Development (OECD)) 2005. године започео је двадесетомесечну студију сврхе, садржаја, квалитета и финансирања отворених образовних ресурса (Hylén, 2006, 1). Завршни извештај објављен је 2007. године, а у њему се закључује да ће отворени образовни ресурси битно утицати на наставни програм, педагошки приступ и оцењивање у оквиру традиционалних институција. Наставни кадар, који уместо "мудраца на позорници" већ постаје "водич са стране", изгубиће и посредничку улогу у обезбеђивању материјала за учење. Потреба за оцењивањем и препознавањем компетенција стечених мимо формалних оквира учења вероватно ће расти. Прилагођавајући се овоме, примарна улога постојећих институција вишег образовања могла би постати оцењивање уместо подучавања (Center for Educational Research and Inovation, 2007). Ове тенденције извештај представља као неизбежне и њима се бави прагматички, а не критички.

На светском конгресу о отвореним образовним ресурсима одржаном 2012. године у седишту Унескоа у Паризу усвојена је декларација (2012 Paris OER Declaration) којом се препоручује да државе у оквиру својих могућности и овлашћења:

1. шире свест о отвореним образовним ресурсима и подстичу њихово коришћење;

[1] UNESCO. "Open educational resources". Промотивни видео, 00:00-00:35 (3:05). Објављено 2012. Преузето 10.08.2016, http://www.unesco.org/archives/multimedia/?s=films_details&id_page=33&id_film=2573

2. омогуће стварање окружења за коришћење информационих и комуникационих технологија (ИКТ);
3. појачају развој стратегија и правила у вези са отвореним образовним ресурсима;
4. промовишу разумевање и коришћење отвореног лиценцирања;
5. подрже изградњу капацитета за одрживи развој квалитетних материјала за учење;
6. подстичу стратешке савезе за отворене ресурсе;
7. охрабрују развој и прилагођавање отворених образовних ресурса на различитим језицима и у различитим културним контекстима;
8. охрабрују истраживање на тему отворених образовних ресурса;
9. олакшају проналажење, преузимање и дељење отворених садржаја;
10. охрабрују отворено лиценцирање образовних материјала финансираних из јавних средстава.

2. Сврха образовања

Питање сврхе образовања често се превиђа у расправама о потребним реформама и расположивим средствима у образовању. Да ли од образовања можемо очекивати развој личних потенцијала и друштвене свести? Шта је основ аутентичног друштвеног доприноса и од каквог је значаја аутентичност сазнајног искуства?

Креативност, индивидуалност, друштвено богатство и култура, као и друштвена свест и одговорност централна су питања како савремене критичке педагогије капиталистичких друштава, тако и социјалистичких реформи спроведених у Совјестком Савезу и другде током 20. века.

Антон Макаренко, један од оснивача совјетске педагогије, сматрао је да је развој личности и карактера основна сврха образовања, и да, према томе, оно мора обухватити све аспекте живота. Он истиче да се пун потенцијал индивидуе и колектива може остварити само уз поштовање и разумевање индивидуалних и колективних склоности, способности и потреба, па су непромењиви обрасци недопустиви у образовању, а средства подучавања морају бити примерена околностима. Захваљујући њему, рад, који мора бити креативан, колективан и продуктиван, добија значајно место у образовању. Овакав рад допушта лични израз, захтева сарадњу и успостављање дисциплине, а спознаја његове друштвене сврхе, односно његовог доприноса економској снази друштва, доноси задовољство, самопоуздање и осећај припадности (Makarenko, 1965).

Кен Робинсон, енглески педагог и аутор, међународни саветник за уметничко образовање при владама, невладиним организацијама, образовним и уметничким телима, наводи четири основне сврхе јавног образовања: економски развој; разумевање сопственог идентитета и идентитета других, као услов за превазилажење сукоба култура, традиција и погледа на свет; развој друштвене свести и одговорности; и развој личних склоности и способности[2].

Отворени образовни ресурси представљају непромењиве обрасце и као средства обуке примерени су специфичним околностима образовних система у којима настају, па самим тим и индивидуалним и колективним склоностима, способностима и потребама које су датим околностима обухваћене и које ови образовни системи препознају. Даље, отворени образовни ресурси доприносе економском развоју развијених земаља, а у складу с тим и ширењу доминантних култура и погледа на свет.

3. Глобално образовање

Глобализација није само економски и политички већ и културни процес, који има пресудни утицај на образовање. Критика глобализације образовања не може се развијати независно од опште критике глобализације као свеобухватног процеса, који подразумева и подстиче надметање међу неједнакима. Највеће економске силе, а уједно и највећи заговорници глобализације, надметање представљају као главни и основни покретач развоја. Примена ове идеје на образовање има далекосежне и трагичне последице. Ученици и студенти надмећу се за место у школама и на универзитетима, школе и универзитети надмећу се за финансијску подршку и опстанак. Наставни програми и организација наставе и наставних активности земаља у развоју прилагођавају се глобалним развојним потребама, које одређују развијене земље. Стандардизовани тестови и рангирање школа треба да обезбеде једнаке критеријуме успеха, док се неједнаки локални услови и различите локалне потребе грубо занемарују.

Програм за међународно вредновање ученика (Programme for Intenrational Student Assessment (PISA)) почео је да се спроводи 2000. године, а

[2] Ken Robinson. “How to change education: From the ground up”. YouTube видео, 01:00-10:00 (24:02). Видео запис предавања одржаног у Краљевском друштву за уметност у Лондону 01.07.2013. године. Постављено “The RSA” 18.07.2013, https://www.youtube.com/watch?v=BEsZOnyQzxQ

сада у њему учествује пола милиона петнаестогодишњака из 65 земаља света и локалних администрација. На основу извештаја Организације за економску сарадњу и развој (OECD, 2014), може се закључити да је друштвено-економски контекст главна одредница успеха земаља учесница на овом тесту, с тим што је, поред укупног улагања у образовање, врло значајно уједначено распоређивање доступних средстава, чиме се постиже једнак квалитет образовања за сву децу. Критичари овог програма, попут учесника ТВ дебате “Is schooling for all a realistic goal?”[3] и бројних потписника писма главном координатору ПИСА програма, Андреасу Шлајхеру (Meyer and Zahedi, 2014, i), истичу да он не узима у обзир читав низ вештина које су од суштинске важности за развој деце. Он по њима превише сужава појам образовања, а успех на овом тесту не одражава стварни квалитет датог образовног система. Не оспоравајући чињеницу да су добри резултати из математике и језика заиста од великог значаја, они наводе друге битне аспекте образовања, међу којима су креативност, самопоуздање и способност интеракције. Уз то, и сама пракса тестирања и рангирања, мада донекле неизбежна и у одређеној мери присутна у свим образовним системима, представља специфичан приступ образовању, који у зависности од предвиђених циљева може бити мање или више продуктиван, али и крајње деструктиван. Огроман притисак на децу земаља успешних на ПИСА тесту наноси озбиљну штету њиховом здрављу (Zhao, 2010), али се у глобалном надметању које програм заступа образовни системи ових земаља ипак намећу као модел за све мање успешне земље (Zhao, 2014).

Сједињене Америчке Државе су 2001. године усвојиле закон No Child Left Behind који је ученике, наставнике и школе приморао на међусобно надметање. Слично ПИСА тестовима, тестови предвиђени овим законом уског су фокуса и недовољно прилагођени специфичним условима одрастања и специфичним склоностима деце, а спроводе се под претњом затварања школа које на њима не постигну адекватан успех. Од школа се тражи да гарантују да ће сва деца усвојити предвиђени минимум вештина у читању, писању и математици, а програми који непосредно не доприносе постизању наведеног минимума губе подршку или бивају поптуно укинути (Leyva, 2009, 10). И мада је промовисан као закон који ће помоћи деци у неповољном положају, бројни критичари

[3] Al Jazeera English. “Is schooling for all a realistic goal?” ТВ дебата, 12:20-15:20 (24:24). Al Jazeera English, 2014. Преузето 12.05.2016, http://www.aljazeera.com/programmes/insidestory/2014/04/schooling-all-realistic-goal-2014410151530585416.html

тврде да су ефекти његове примене управо супротни. А ево шта деца кажу о овим тестовима:

„Не могу да издржим, стварно не могу. Толико се стидим". "Цео свет ми је на раменима. Ово ће ући у мој досије, на основу којег уписујем колеџ, а колеџ значи добар посао". „Руке су ми се знојиле, готово сам заплакала, и имала сам чудан осећај у стомаку. Надам се да се никад више нећу тако осећати, посебно не због неког теста на којем не могу да покажем шта сам све дивно урадила током године". „Зашто пресуђивати на основу једног теста? Зашто не на основу напретка током године"? „Овај тест не показује оно што радите најбоље, као на пример писање, мени омиљено. Сваког дана ја сам све јача, показује ли то? Не..." „Осећам да не би требало да смо оволико нервозни овако мали".[4]

Разумевање појаве отворених образовних ресурса захтева бављење не само технолошким околностима њиховог настанка и условима и могућностима за њихов даљи развој, већ и друштвено-економским и идеолошким оквирима у којима се они представљају наизменично као допуна институционалним програмима, решење за недостатак институционалних капацитета, али и као радикални изазов доминацији институција (Knox, 2013, 4). Њихова појава временски се подудара са првим ПИСА тестом, усвајањем поменутог закона у Сједињеним Америчким Државама и Светским образовним форумом одржаним 2000. године у Дакару, о чему више у наставку. Све ове иницијативе посвећене су увећавању „људског капитала"[5,6] (Kwon, 2009) уместо остваравању људског потенцијала, и образовању „глобално конкурентне радне снаге" (Spellings, 2012), често на штету развоја локално релевантних компетенција.

[4] Lerone D. Wilson. No Child Left Behind. Документарни филм. Boondoggle Media, 2004. YouTube видео, 00:00-01:45 (56:02). Објављено „Boondoggle" 06.04.2011. Преузето 15.06.2016, https://www.youtube.com/watch?v=yiGN7kVyeaM

[5] UN Deapartment of Economic and Social Affairs (DESA). „Linking education to human capital development". Објављено 21.03.2011. Преузето 21.05.2016, http://www.un.org/en/development/desa/news/ecosoc/linking_education.html

[6] UN Deapartment of Economic and Social Affairs (DESA). „Human and natural capital as important as financial capital". Објављено 17.05.2012. Преузето 21.05.2016, http://www.un.org/en/development/desa/news/financing/new-paradigm-job-creation.html

У капиталистистичком друштву систем образовања подређен је интересима капитала и има за циљ легитимизацију и репродукцију постојећих друштвених односа (Bukharin, 2001). Друштвена вредност образовања, као и сваког другог друштвеног програма, губи значај, а оно постаје питање личних жеља и личне користи, чиме се објашњава искључивање система образовања из домена друштвене организације и одговорности. Свако ко очекује извесну корист од сопственог образовања слободан је да према својим очекивањима у њега и улаже, али је и обавезан да сам сноси ризик тог улога. Преокупираност сопственим амбицијама и успехом ученике и студенте наводи на надметање, на штету сарадње и бриге за колективни напредак и образовно окружење (Leyva, 2009, 10) (Saunders, 2010). Отворени образовни ресурси ослањају се на самоинцијативу и самоусмереност својих корисника, јер не нуде динамично образовно окружење које подржава стварање аутентичног колектива. Упркос овоме, они се промовишу као најбољи образовни ресурси: њих креира најбољи наставни кадар, подржани су најбољим технологијама и доступни свима који за себе желе најбоље[7] (Knox, 2013, 4).

Велико интересовање за отворене образовне ресурсе одраз је значаја и утицаја институција које их креирају, али и основ ширења њихове интелектуалне доминације. Отворене образовне ресурсе у највећој мери креирају елитне институције развијених земаља, при чему се све остале институције лако своде на пасивне кориснике доступних ресурса, без доприноса или са малим доприносом глобалној размени и протоку знања. Постојећи поредак, односно економска и технолошка супериорност елитних институција, на овај начин се утврђује и јача. Уз одговарајуће акредитације овакав приступ образовању могао би учинити многе традиционалне институције редундантнима. Није тешко замислити да би нас овај тренд, уз краткорочни привид флексибилности и смањења трошкова, у догледно време могао довести до глобалне монополизације образовања.

Отворени образовни ресурси још увек су недовољно дефинисани, па се мора водити рачуна о евентуалним препрекама њиховом слободном коришћењу. Формати који захтевају заштићен софтвер за приступ и модификацију, као и технички ограничене платформе за дистрибуцију

[7] UNESCO. „Open educational resources”. Промотивни видео, 00:43-01:03 (3:05). Објављено 2012. Преузето 10.08.2016, http://www.unesco.org/archives/multimedia/?s=films_details&id_page=33&id_film=2573

садржаја битно ограничавају употребљивост отворених образовних ресурса. Отворени образовни ресурси ретко оправдавају свој назив ако се узме у обзир и конфузија која постоји по питању одговарајућих отворених лиценци: за отворене образовне ресурсе често се узимају и бесплатни образовни ресурси који не дозвољавају прераду, што значи да их без одговарајућих дозвола није могуће превести на друге језике и прилагодити различитим локалним контекстима (Hoel, 2014), као у случају већине масивних отворених онлајн курсева (massive open online course (MOOC)).

4. Глобална неједнакост

Обезбеђивање основног образовања за сву децу света до 2015. године један је од основних Миленијумских развојних циљева Уједињених нација. Према последњим подацима, 263 милиона деце узраста од 6 до 17 година не похађа школу, при чему 93,3 милиона њих живи у Субсахарској Африци, а 100,8 милиона у Јужној Азији[8]. Према извештају који су Уједињене нације објавиле 2014. године, 125 милиона деце у свету није усвојило основне вештине читања и писања чак ни после четири године основношколског образовања, што доводи у питање и квалитет наставе и наставног кадра (UNESCO, 2014, i). На светском образовном форуму у Дакару 2000. године постављено је шест главних циљева које је требало остварити до 2015. године (UNESCO, 2000, 15–17):

1. ширити и унапредити свеобухватну рану бригу и образовање, посебно за најрањивију и најугроженију децу;
2. обезбедити да до 2015. године сва деца, а посебно девојчице, деца у тешким околностима и деца из етнички мањинских група, имају приступ бесплатном и обавезном основном образовању доброг квалитета, као и да га заврше;
3. обезбедити да се изађе у сусрет сазнајним потребама свих младих и одраслих кроз равноправан приступ одговарајућим програмима учења и обуке;
4. до 2015. године за 50% повећати писменост одраслих, посебно жена, и осигурати равноправан приступ основном и даљем образовању за све одрасле;

[8] UNESCO Institute for Statistics. “263 million children and youth are out of school”. Објављено 15.07.2016. Преузето 22.08.2016, http://uis.unesco.org/en/news/263-million-children-and-youth-are-out-school

5. елиминисати велике родне разлике у основном и средњем образовању до 2005. године и постићи родну једнакост у образовању до 2015. године;
6. унапредити све аспекте квалитета образовања и осигурати њихов висок стандард тако да сви постижу признате и мерљиве резултате, посебно у писмености, рачунању и основним животним вештинама.

Унеско упозорава да ови циљеви не могу бити остварени без значајане промене у државном приступу обезбеђивању образовања. Недостатак политичке воље на националном нивоу праћен је недостатком политичке воље међународне заједнице. Када су ови циљеви постављени, 2000. године, постигнут је договор да ће свака земља која буде имала план имати и средства за остваривање тог плана, што се није десило у пракси[9].

Глобално, у зонама конфликта, број деце која не похађају школу достигао је ове године 61,9 милиона.[10] Број сиријске деце која не похађају школу, било да се налазе у ратом захваћеној Сирији или у избеглиштву, тренутно је два милиона и свакодневно расте.[11] У школама Газе у септембру 2014. године, после двомесечног бомбардовања од стране Израела, школска година почела је са закашњењем и то не наставом него колективном терапијом (Weibel, 2014).

Ратови спречавају развој, а разарањем инфраструктуре и од развијених земаља могу створити земље у развоју (Zolnikov, 2013). Инфраструктурални проблеми отежавају борбу са болестима и природним катастрофама, а корупција и организовани криминал отежавају решавање инфраструктуралних проблема.[12] У случају Хаитија, опоравак од последица земљотреса и епидемије колере,

[9] Al Jazeera English. „Is schooling for all a realistic goal?”. ТВ дебата, 02:55-03:15 (24:24). Al Jazeera English, 2014. Преузето 12.05.2016, http://www.aljazeera.com/programmes/insidestory/2014/04/schooling-all-realistic-goal-2014410151530585416.html

[10] UNESCO Institute for Statistics. „263 million children and youth are out of school”. Објављено 15.07.2016. Преузето 22.08.2016, http://uis.unesco.org/en/news/263-million-children-and-youth-are-out-school

[11] Save the Children. „Children of Syria”. Објављено 2016. Преузето 15.08.2016, http://www.savethechildren.org/site/c.8rKLIXMGIpI4E/b.7998857/k.D075/Syria.htm

[12] UN Interregional Crime and Justice Research Institute. „Organized crime and corruption”. Преузето 27.07.2016, http://www.unicri.it/topics/organized_crime_corruption

којима је погођен 2010. године, отежан је политичком нестабилношћу, неодговорношћу међународних организација (Pilkington, 2016) и лошом координацијом бројних добротворних организација и иницијатива (Knox, 2015). Процењује се да на Хаитију око 50% деце никада неће ући у учионицу.[13] Школе земаља у развоју често немају струју, пијаћу воду и основне хигијенске услове. Оне школе које имају какву-такву струју ретко имају и интернет. А ако и имају интернет, недостаје им наставни кадар за спровођење информатичке наставе (Cave, 2013)(Mungai, 2011). Наставници ових земаља приморани су да раде са великим бројем ученика без адекватних средстава и потребне подршке, чиме су могућности за увођење новина у наставне програме значајно ограничене, а превиђање ових ограничења може за резултат имати отуђивање наставника и урушавање ионако крхких образовних система.[14]

У овако екстремним околностима питања иновација у образовању, у виду савремених метода и технологија, и сходно томе одговарајуће обуке наставника, чине се апсурднима. Па ипак, постоје школе у које рачунари стижу пре редовног напајања електричном енергијом. Пример су мобилне соларне учионице у Уганди[15] и дигитална села у Јужној Африци (Cave, 2013). Ово су донаторски пројекти, али како и развој инфраструктуре у овим земљама углавном зависи од донација, редослед приоритета свакако је збуњујући. Добротворне и невладине организације које спроводе ове пројекте, уз подршку великих корпорација попут Интела и Мајкрософта, као свој циљ наводе „извлачење Африке из замке сиромаштва и обучавање следеће

[13] Al Jazeera English. „Is schooling for all a realistic goal?". ТВ дебата, 19:55-20:00 (24:24). Al Jazeera English, 2014. Преузето 12.05.2016, http://www.aljazeera.com/programmes/insidestory/2014/04/schooling-all-realistic-goal-2014410151530585416.html

[14] Al Jazeera English. „Is schooling for all a realistic goal?". ТВ дебата, 17:10-17:15 (24:24). Al Jazeera English, 2014. Преузето 12.05.2016, http://www.aljazeera.com/programmes/insidestory/2014/04/schooling-all-realistic-goal-2014410151530585416.html

[15] Maendeleo Foundation. „Mobile solar computer classrooms". Објављено 2015. Преузето 23.05.2016, http://maendeleofoundation.org/projects/mobile-solar-computer-classrooms

генерације за рад у глобалном окружењу”[16], али и „развој бизниса у области информационо-комуникационих технологија”[17].

Многа деца земаља у развоју приморана су да на разне начине помажу својим родитељима у прехрањивању породице, радећи послове који угрожавају њихово здравље или помажући у кућним пословима и у одгоју млађе деце док родитељи раде[18]. Овој деци чак и бесплатно образовање остаје недоступно. У земљама у развоју велике ИКТ компаније истовремено експлоатишу дечју радну снагу (Russell, 2016) и спонзоришу отворене образовне ресурсе путем којих се промовишу[19].

Ако се има у виду да ИКТ индустрија, као и свака друга, упошљавањем јефтине радне снаге постиже већи профит и нижу цену производа и услуга, не треба да изненади овакво интересовање за оспособљавање нових генерација ИКТ радника у земљама у развоју. Минимална цена рада у Кини тренутно је око 9 пута нижа од законом предвиђеног минумума у Америци[20][21], у Индији око 20 пута, а у Уганди 725 пута[22]. Пословање светских гиганата ИКТ индустрије у Кини и Индији добро илуструје могући развојни пут земаља које у овом сектору траже своју развојну шансу[23] (Chakrabortty, 2013)(Garside and Arthur, 2013).

---

[16] Computers 4 Africa. „About Computers 4 Africa”. Преузето 23.05.2016, http://www.computers4africa.org.uk/about/index.php

[17] Maendeleo Foundation. „About us”. Преузето 23.05.2016, http://maendeleofoundation.org/our-story/

[18] UNICEF. „Child labour”. UNICEF, 11.06.2015. Преузето 20.05.2016, https://www.unicef.org/protection/57929_child_labour.html

[19] OER Africa. „Who we are”. Преузето 20.08.2016, http://www.oerafrica.org/about-us/who-we-are

[20] „Minimum wages”. WageIndicator, 2016. Преузето 23.07.2016, http://www.wageindicator.org/main/salary/minimum-wage

[21] United States Department of Labor. „Minimum Wage”. United States Department of Labor. Преузето 23.07.20016, https://www.dol.gov/general/topic/wages/minimumwage

[22] WageIndicator. „Minimum wages”. WageIndicator, 2016. Преузето 23.07.2016, http://www.wageindicator.org/main/salary/minimum-wage

[23] Al Jazeera English. „India: High tech mirage”. Документарни програм, 00:00-00:35 (24:59). Al Jazeera English, 2014. Преузето 25.05.2016, http://www.aljazeera.com/programmes/peopleandpower/2014/03/india-high-tech-mirage-201431985356939298.html

5. Закључак

Отворени образовни ресурси промовишу се као посебно значајни и корисни за оне који желе да уче, а немају могућност да се укључе у традиционалне моделе учења. Оваквим потенцијалним корисницима нуди се еманципација и ослобађање од угњетавања и лоших услова живота (Knox, 2013, 7), иако су за крајње лоше услове живота и угњетавање широм света неретко одговорни управо спонзори институција које креирају отворене образовне ресурсе[24,25] (Washburn, 2010, 5) (Palomino, 2013) (McClenaghan, 2015) (Burgis, 2015).

Неједнако образовање последица је, а не узрок основних економских неједнакости. Тврдња да образовање нуди излаз из сиромаштва свеприсутна је како у промоцији отворених образовних ресурса тако и у обухватнијим међународним кампањама и апелима. Најсиромашнијима, међутим, оно нуди само наду у потпуном безнађу. Дечак на Хаитију своје похађање школе објашњава на следећи начин: „Мама и тата шаљу ме у школу да учим тако да кад порастем могу да имам добар живот и добар ауто, као и свако други", а његова учитељица истиче: „Без школе, нема живота. Ако постанеш председник, то је зато што си ишао у школу".[26].

Образовним системима којима недостају финансијска средства не може се помоћи додатним финансирањем најбоље финансираних образовних система. Материјали за учење и подучавање само су део сложеног процеса образовања, па њихова доступност само делимично доприноси доступности образовања. Сваки допринос доступности образовања заслужује сразмерну хвалу и подршку, али кампање, апели и пропагандни материјали који позивају на подршку отвореним

[24] Multinational Monitor. "BP: A legacy of apartheid, pollution and exploitation". Multinational Monitor, 1992. Преузето 15.06.2016, http://www.multinationalmonitor.org/hyper/issues/1992/11/mm1192_11.html

[25] Al Jazeera English. „The Secret of the Seven Sisters". Documentary series in four parts with an accompanying description. Al Jazeera English, 26 April 2013. Преузето 28.07.2016, http://www.aljazeera.com/programmes/specialseries/2013/04/201344105231487582.html. YouTube видео, 3:10:02. Постављено „Rebaz Zedbagi" 25.05.2014. Преузето 28.07.2016, https://www.youtube.com/watch?v=XtYOjMmEMeg

[26] Al Jazeera English. „Is schooling for all a realistic goal?". ТВ дебата, 19:10-19:45 (24:24). Al Jazeera English, 2014. Преузето 12.05.2016, http://www.aljazeera.com/programmes/insidestory/2014/04/schooling-all-realistic-goal-2014410151530585416.html

образовним ресурсима доступност поистовећују са примењивошћу, стварајући погрешну представу о могућем доприносу отворених образовних ресурса унапређењу глобалног образовања (Bates, 2011) (Crissinger, 2015). Отворени образовни ресурси не нуде него подразумевају адекватна решења за инфраструктуралне, кадровске и бројне друге проблеме образовних система како развијених земаља тако и земаља у развоју, што води превиђању а не решавању ових проблема. Највећу штету, наравно, трпе земље са највећим проблемима.

Литература

Bates, Tony. “OERs: The good, the bad and the ugly”. *Online Learning and Distance Education Resources (blog)* (06.02.2011). Преузето 14.07.2016, http://www.tonybates.ca/2011/02/06/oers-the-good-the-bad-and-the-ugly

Nikolai Bukharin and Evgenii Preobrazhensky. *The ABC of Communism*. Marxists Internet Archive, 1965

Burgis, Tom. “Big Oil’s sleazy Africa secrets: How American companies and super-rich exploit natural resources”. *Salon* (06.04.2015). Преузето 15.06.2016, http://www.salon.com/2015/04/06/big_oils_sleazy_africa_secrets_how_american_companies_and_super_rich_exploit_natural_resources

Cave, Kathryn. “South Africa: Why have all the rural tech projects failed?”. *IDG Connect*, 21.06.2013. Преузето 25.05.2016, http://www.idgconnect.com/blog-abstract/2292/south-africa-why-have-all-rural-tech-projects-failed

Center for Educational Research and Inovation (CERI). *Giving Knowledge for Free: The Emergence of Open Educational Resources*. Paris: OECD, 2007. Преузето 15.06.2016, https://www.oecd.org/edu/ceri/38654317.pdf

Chakrabortty, Aditya. “The woman who nearly died making your iPad”. *The Guardian* no. 25(25.08.2013). Преузето 25.05.2016, https://www.theguardian.com/commentisfree/2013/aug/05/woman-nearly-died-making-ipad

Crissinger, Sarah. “A critical take on OER practices: Interrogating commercialization, colonialism and content”. *In the Library with the Lead Pipe (blog)* (21.10.2015). Преузето 14.07.2016, http://www.inthelibrarywiththeleadpipe.org/2015/a-critical-

take-on-oer-practices-interrogating-commercialization-colonialism-and-content

Garside, Juliette and Charles Arthur. "Workers' rights 'flouted' at Apple's iPhone factory in China". *The Guardian* no. 5(05.09.2013). Преузето 25.05.2016, https://www.theguardian.com/technology/2013/sep/05/workers-rights-flouted-apple-iphone-plant

Hoel, Tore. "Adopting OER for less used languages: We need hard talk on tools and infrastructure!". *Nordic OER*, 11.04.2014. Архивирано Internet Archive Way Back Machine 11.03.2015. Преузето 11.08.2016, https://web.archive.org/web/20150311234754/http://nordicoer.org/adopting-oer-less-used-languages-need-hard-talk-tools-infrastructure

Hylén, Jan. "Mapping users and producers of open educational resources". UNESCO, November 2006. Преузето 25.05.2016, http://www.unesco.org/iiep/virtualuniversity/media/forum/iiep_oecd_oer_forum_note1.pdf

Knox, Jeremy. "Five critiques of the open educational resources movement". *Teaching in Higher Education* Vol. 18, no. 8(2013): 821–832. Преузето 02.08.2016, http://www.academia.edu/2651447/Knox_J._2013_Five_Critiques_of_the_Open_Educational_Resources_Movement._Teaching_in_Higher_Education

Knox, Richard. "5 years after Haiti's earthquake, where did the $13.5 billion go?". *NPR*, 12.01.2015. Преузето 20.05.2016, http://www.npr.org/sections/goatsandsoda/2015/01/12/376138864/5-years-after-haiti-s-earthquake-why-aren-t-things-better

Kwon, Dae-Bong. "Human capital and its measurement". In *3rd OECD World Forum on "Statistics, Knowledge and Policy": Charting Progress, Building Visions, Improving Life. Busan, South Korea, 27-30 October 2009*, 2009. Преузето 01.08.2016, http://www.oecd.org/site/progresskorea/44109779.pdf

Leyva, Rodolfo. "No Child Left Behind: A neoliberal repackaging of social darwinism". *Journal for Critical Education Policy Studies* Vol. 7, no. 1(2009): 365–381. Преузето 25.07.2016, http://www.jceps.com/wp-content/uploads/PDFs/07-1-15.pdf

Makarenko, Anton S. *Problems of Soviet School Education*. Moscow: Progess Publishers, 1965

McClenaghan, Meave. "Investigation: Top universities take £134m from fossil fuel giants despite divestment drive". *Greenpeace Energydesk*, 23.10.2015. Преузето 15.06.2016, http://energydesk.greenpeace.

org/2015/10/23/data-top-universities-take-134m-from-fossil-fuel-giants-despite-divestment-drive

Meyer, Heinz-Dieter and Katie Zahedi. “Open letter to Andreas Schleicher”. *Global Policy Journal* (05.05.2014). Преузето 20.05.2016, http://www.globalpolicyjournal.com/blog/05/05/2014/open-letter-andreas-schleicher-oecd-paris

Mungai, Martin. “12 challenges facing computer education in Kenyan schools”. *ICT Works*, 12.09.2011. Преузето 29.07.2016, http://www.ictworks.org/2011/09/12/12-challenges-facing-computer-education-kenyan-schools

OECD. “How is equality in resource allocation related to student performance?”. *PISA in Focus* no. 44(2014). Преузето 17.05.2016, https://www.oecd.org/pisa/pisaproducts/pisainfocus/pisa-in-focus-n44-(eng)-final.pdf

Palomino, Joaquin. “The university of private enterprise”, *East Bay Express* (10.04.2013). Преузето 15.06.2016, http://www.eastbayexpress.com/oakland/the-university-of-private-enterprise/Content?oid=3518686

Pilkington, Ed. “UN makes first public admission of blame for Haiti cholera outbreak”. *The Guardian* no. 18(18.08.2016). Преузето 20.05.2016, https://www.theguardian.com/world/2016/aug/18/un-public-admission-haiti-cholera-outbreak

Russell, Jon. “Apple, Microsoft, Samsung and other tech firms implicated in child labour report”. *TechCrunch* (19.01.2016). Преузето 27.05.2016, https://techcrunch.com/2016/01/19/apple-microsoft-samsung-and-other-tech-firms-implicated-in-child-labor-report

Saunders, Daniel B. “Neoliberal ideology and public higher education in the United States”. *Journal for Critical Education Policy Studies* Vol. 8, no. 1(2010): 41–77. Преузето 25.07.2016, http://www.jceps.com/wp-content/uploads/PDFs/08-1-02.pdf

Spellings, Margaret. “Building a globally competitive workforce”. US Chamber of Commerce Foundation, 31.05.2012. Преузето 04.08.2016, https://www.uschamberfoundation.org/newsletter-article/building-globally-competitive-workforce

UNESCO. “The Dakar Framework for Action: Education for All: Meeting Our Collective Commitments”, In *World Education Forum, Dakar, Senegal, 26–28 April*. UNESCO, 2000. Преузето 15.05.2016, http://unesdoc.unesco.org/images/0012/001211/121147e.pdf

UNESCO. “Teaching and Learning: Achieving Quality for All (Education for All Global Monitoring Report, 2013/4)”. Techrep. UNESCO, 2014.

Преузето 15.05.2016, http://en.unesco.org/gem-report/report/2014/teaching-and-learning-achieving-quality-all

Washburn, Jennifer. "Big Oil Goes to College: An Analysis of 10 Research Collaboration Contracts between Leading Energy Companies and Major U.S. Universities". Techrep. Center for American Progress, 2010. Преузето 15.06.2016, https://www.americanprogress.org/wp-content/uploads/issues/2010/10/pdf/big_oil_lf.pdf

Weibel, Catherine. "In Gaza, children go back to school after a devastating summer". UNICEF, 2014. Преузето 10.06.2016, http://www.unicef.org/education/oPt_75921.html

Zhao, Yong. "A true wake-up call for Arne Duncan". *Education in the Age of Globalization (blog)* (10.12.2010). Преузето 15.05.2016, http://zhaolearning.com/2010/12/10/a-true-wake-up-call-for-arne-duncan-the-real-reason-behind-chinese-students-top-pisa-performance

Zhao, Yong. "How does PISA put the world at risk (part 1): Romanticizing misery". *Education in the Age of Globalization (blog)* (09.03.2014). Преузето 15.05.2016, http://zhaolearning.com/2014/03/09/how-does-pisa-put-the-world-at-risk-part-1-romanticizing-misery

Zolnikov, Tara Rava. "From developed to developing country: Water and war in Iraq". *Johns Hopkins Water Magazine* (31.05.2013). Преузето 05.07.2016, http://water.jhu.edu/index.php/magazine/from-developed-to-developing-country-water-and-war-in-iraq

Сврставање појмова на континуалној скали позитивно-негативно према поларитету осећања засновано на емотиконима

УДК 811.163.41’322.2

Михаило Шкорић
miks@tesla.rcub.bg.ac.rs
Универзитет у Београду

САЖЕТАК: Циљ овог рада је скретање пажње на могућност коришћења текста на вебу обележеног емотиконима и другим одређивачким нискама у сентименталној анализи језички независног корпуса. Он уводи неколико новина у постојећу истраживачку окосницу и тестира њихову успешност. Такође, приказује софтверски алат специјално направљен у ту сврху, објашњава систем његовог рада у изради базе података са појмовима који одражавају одређени степен осећања и даје упуство за његово коришћење. Осим тога, приказује и софтверски алат за тестирање базе података и даје примере анализе добијених резултата.

КЉУЧНЕ РЕЧИ: истраживање података, поларитет осећања, екстракција информација, емотикони, текст на вебу.

РАД ПРИМЉЕН: 24. јануар 2017.
РАД ПРИХВАЋЕН: 25. март 2017.

1. Увод

Приликом израде софтвера за разумевање природног језика, широко су прихваћена два приступа:

– Софтвер који нема дубоко разумевање значења текста, већ само граматике језика којим је он написан, што омогућава ширу примену.

– Софтвер који има дубље разумевање самог значења текста, често ограничен на једну или само неколико области које покушава да разуме и претежно се користи за класификацију текстова.

Системи засновани на анализи осећања додељују тексту вредности над параметрима осећања, где већи број параметара значи много већу комплексност дефинисања.

Под претпоставком да се приликом сравњивања свих могућих параметара на поларитет осећања не долази до великог губитка информација, и да машина може да одлучи између више избора мерењем једног јединог параметра, задатак се своди на одређивање тога шта је позитивно, а шта негативно.

У циљу што економичнијег прављења интелигентног система, не треба да се занемари ниједан ресурс који стоји на располагању и који може потенцијално да буде од користи. Основна идеја овог истраживања је коришћење метода истраживања података за проналажење метаподатака – одређивача, које корисници друштвених мрежа ненамерно остављају у својим порукама (у облику емотикона или језички-универзалних устаљених фраза) и додељивање вредности позитивности појмовима у скупу у којем се они налазе. Како су одређивачи језички независни, и овај систем би био језички независтан. Могао би, уколико се испостави као валидан, да омогући коришћење метода машинског учења над огромним корпусом текстова унапред обележених одређивачима.

1.1 Осврт на пређашња слична истраживања

Године 2005. објављена је серија експеримената са класификацијом расположења постова на интернет блоговима, која је послужила као основа за многа будућа истраживања (Mishne, 2005). Ови експерименти састојали су се од евиденције појмова који су се појављивали у постовима за које су сами аутори тврдили да су одражавали одређено расположење док су их писали. Направљени су индекси појмова и израчуната је њихова фреквенција у постовима који су поистовећивани са једним од девет различитих расположења: забављеност, умор, срећа, радост, досада, осећај успеха, поспаност, задовољство и узбуђење. Нови постови су затим тестирани на фреквенцију речи како би се утврдило расположење аутора који их је писао. Резултати су упоређени са људском проценом истих постова и закључено је да машина процењује расположење аутора само мало лошије од човека.

На Универизитету у Токију је 2009. године објављено истраживање у оквиру кога се у тексту анализирало девет расположења, коришћењем комплексних коначних аутомата који препознају граматичке структуре текста (Neviarouskaya et. al., 2009). Исте године истраживачи са Универзитета Станфорд у Калифорнији представили су нови систем анализе веб-постова коришћењем алгоритама који су тренирани да препознају емотиконе и који су додељивали позитивну или негативну вредност расположења порукама на *Twitter* платоформи (Go et. al., 2009). Циљ истраживања био је да се направи систем за класификацију постова, како би потрошачи могли да истраже став претходних купаца пре куповине неког производа. Неколико различитих алгоритама за машинско учење тренирано је са осам емотикона (пет позитивних и три негативна), а резултати су показали да је њихова прецизност изнад 80% у погађању расположења у постовима.

Истраживачи са Хебрејског универзитета у Јерусалиму су 2010. године спровели још једно слично истраживање расположења израженог у постовима на платформи *Twitter*, узимајући у обзир, поред 15 емотикона, и 50 тагова,[1] што је њихов оригинални допринос (Davidov et. al., 2010). Алгоритми који су тренирани на таговима, такође су показили успешност у препознавању расположења поста.

У поменутим истраживањима емотикони се у тексту третирају као ниске карактера експлицитног значења. Другачији приступ предложили су 2010. године истраживачи са Хокаидо универзитета у Сапору. Незадовољни дотадашњим базама емотикона и њиховим вредностима, бавили су се пре свега тиме како да њихову вредност утврде што тачније. Идеја је била да се емотикони третирају као структура састављена од засебних елемената који представљају очи и уста. Композитни делови посебно су обрађивани како би се израчунала њихова вредност. Када су направили своју базу у њој су се емотикони оцењивали према десет могућих осећања: бес, негодовање, узбуђење, страх, допадање, срећа, олакшање, срамота, туга и изненађење (Ptaszynski et. al., 2010). Ово истраживање је касније проширено, па су 2011. године објавили рад на тему истраживања емотикона у којем су емотикони дефинисани као делови природног језика, те је предложено да њихово истраживање буде укључено у истраживања природног језика (Ptaszynski et. al., 2011).

[1] Корисници платформе *Twitter* имају опцију да своје постове додатно обележе таговима како би постови који говоре о некој теми могли лакше да буду пронађени.

У оквиру овог истраживања спроведена је анкета која је показала да су емотикони други по реду носиоци осећања у реченици после самих лексичких израза.

1.2 Основне информације о експерименту

Овај експеримент има циљ да испроба нови приступ истраживању текста екстракцијом емотикона на нови начин, комбинацијом следеће три идеје:

- Емотиконима који ће се користити у експерименту неће се доделити дискретне вредности попут позитивно или негативно већ ће им бити додељена вредност на скали од најпозитивније до најнегативније. Ово ће се одразити и на појмове чија ће вредност, такође, бити додељена на овој скали.
- Користиће се искључиво универзалне одређивачке ниске независне од специфичног језика. Циљ је да се направи у потпуности језички независтан систем што би умногоме проширило могући корпус учења.
- Уместо засебних порука, попут оних на платформи *Twitter*, користиће се поруке које су део конверзације. Циљ овога је да се тестира могућност деловања одређивача не само на поруку у којој се налазе, већ и на поруке у непосредној близини.

Додатни резултат овог експеримента су и два софтверска алата, специјално направљена ради обављања експеримента. Ови програми ће омогућити будућим истраживачима да тестирају сличне идеје, а детаљно објашњење рада софтвера може помоћи и при изради потпуно новог, бољег и потпунијег алата.

Основа идеја овог експеримента јесте да докаже да је могуће:

- језички независно сачинити инвертовани индекс појмова над корпусом текстова који садржи скуп познатих одређивача.
- помоћу тих одређивача аутоматски доделити појмовима вредности на скали позитивно-негативно, који одражава став људи о тим појмовима.

Експеримент се састоји од два дела:

- аутоматска израда базе података која садржи инвертовани индекс појмова и њихових вредности, коришћењем софтверског алата специјално направљеног за овај експеримент.

– тестирање вредности појмова из базе података упоређивањем са исказима људи и коришћењем софтверског алата за тестирање који је наменски направљен за базе података које су продукт првог дела експеримента.

2. Припрема за израду базе података

База података, продукт овог експеримента, окупља на једном месту појмове (речи које се јављају у разговорима) и вредности расположења које они одражавају (уколико постоје) у бројчаном облику. Вредности су израчунате у односу на то у којој близини у тексту у односу на посматрани појам се јављају одређивачи вредности, који је облик тих одређивача и која је њихова вредност. Одређивачи вредности могу бити емотикони или устаљене фразе које су у разговору јављају, а које по природи нису универзалног значења и одражавају позитиван или негативан став, замењујући тако у писаном тексту изразе лица и интонацију.

Интензитет вредности одређивача директно утиче на интензитет вредности коју он преноси појму. Такође, што је одређивач ближи појму, то ће се његова вредност више одразити на вредност самог појма.

Како би формирање овакве базе било успешно и њен коначни исход задовољавајући, потребно је задовољити три предуслова:

– корпус прикупљених текстова мора да буде организован на одређен начин;
– прикупљени текстови и поруке морају да садрже одређиваче који би допринели додели вредности појмовима који се уз њих налазе;
– одређивачи морају имати унапред одређене вредности.

У наставку текста биће детаљно објашњено како је направљена база података за ово истраживање, од прикупљања самог корпуса тесктова до извоза коначне и потпуне базе података, која се онда може користити на више начина.

2.1 Прикупљање корпуса текстова

Полазна идеја је била да се база као резултат овог истраживања заснива на корпусу текстова који садрже преписку и у оквиру ње одређиваче да изразе позитивне или негативне ставове саговорника.

За овај експеримнет коришћене су датотеке које садрже историју дописивања корисника на друштвеној мрежи *Facebook*, за чију је припремну обраду припремљена одговарајућа XSL трансформација (објашњено у одељку 3.1). Неколицина добровољаца је ове датотеке преузела са званичне веб-странице[2] и проследила их аутору овог рада како би биле укључене у корпус и употребљене у истраживању.

Прикупљено је и коришћено шест датотека (од шест корисника) величине од 1,85MB до 167,09MB. Оне заједно садрже 3.884 разговора 2019 различитих корисника, и 1.843.826 порука које су ти корисници међусобно разменили. Садржај ових порука послужио је приликом прављења базе за ово истраживање. Тиме су испуњена прва два претходно наведена предуслова.

2.2 Додељивање вредности одређивачима

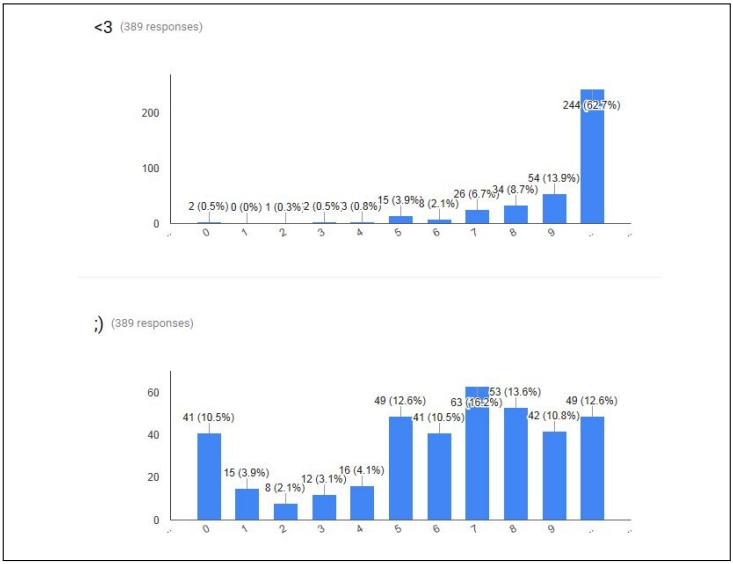
Вредности одређивача добијене су методом анкете која је спроведена на интернет платформи *Google forms*, којој су људи могли да приступе путем хиперлинка објављеног на друштвеним мрежама или прослеђеног од стране других учесника. Учесници анкете су добили задатак да задатим емотиконима и фразама доделе вредност између 0 и 10, где 0 представља највећи интензитет негативног расположења, а 10 највећи интензитет позитивног расположења. Било им је наложено да се при оцењивању воде тиме како одређивач одражава њихово осећање када га шаљу или како га тумаче у примљеној поруци. Учесници анкете су такође имали могућност да сами предложе додатне одређиваче за које су сматрали да су релевантни и предложе њихове вредности.

У периоду од 30 дана 389 учесника анкете оценило је 19 предложених одређивача и предложило додатна 22, од којих је 9 прихваћено.[3] Резултати су се разликовали од одређивача до одређивача – негде су резултати били уједначенији, док негде нису (слика 1).

[2] Друштвена мрежа *Facebook* омогућава свим својим корисницима да путем странице подешавања на свом профилу преузму датотеку која садржи све њихове тренутне мултимедијалне датотеке, као и историју ћаскања у облику јединствене ZIP датотеке.

[3] Укупно 143 корисника су предложила неки одређивач, а потребан број предлагача да би се нови одређивач уврстио био је 48, тј. више од једне трећине предлагача. Уколико је довољан број корисника предложио неки одређивач његова вредност је израчуната као средња вредност свих предложених вредности и уврштен је у истраживање.

По завршетку испитивања, за вредност сваког одређивача узета је аритметичка средина скупа вредности које су испитаници проследили (или предложили за 9 накнадно додатих одређивача).



Слика 1. Пример различите усаглашености корисника над значењем одређивача *<3* (велика сагласност – преко 60% испитаника се слаже око једног одговора) и значењем одређивача *;)* (лоша сагласност – чак пет најчешће одабраних одговора у распону од само 5.7%).

Како би вредности боље осликале оно што треба да представљају – тачку на скали негативно-позитивно – оне су пресликане из скупа $(0, 10)$ на скуп $(-1, 1)$, што је учињено применом формуле:

$$x = (x - 5)/5 \tag{1}$$

Тако је вредност која одговара негативном расположењу највећег интензитета постала -1, док је вредност која одговара позитивном расположењу највећег интензитета постала 1. Израчунат је, такође, универзални фактор корекције којим су све вредности помножене. За овај фактор узет је количник највеће вредности 1 и вредности

одређивача са највећим позитивним интензитетом $0,81$, тако да одређивач *<3*, који према учесницима у анкети одражава позитивно расположење највећег интензитета, добио максималну вредност 1. Тиме што су свим одређивачима додељене одговарајуће вредности испуњен је трећи услов за формирање базе.

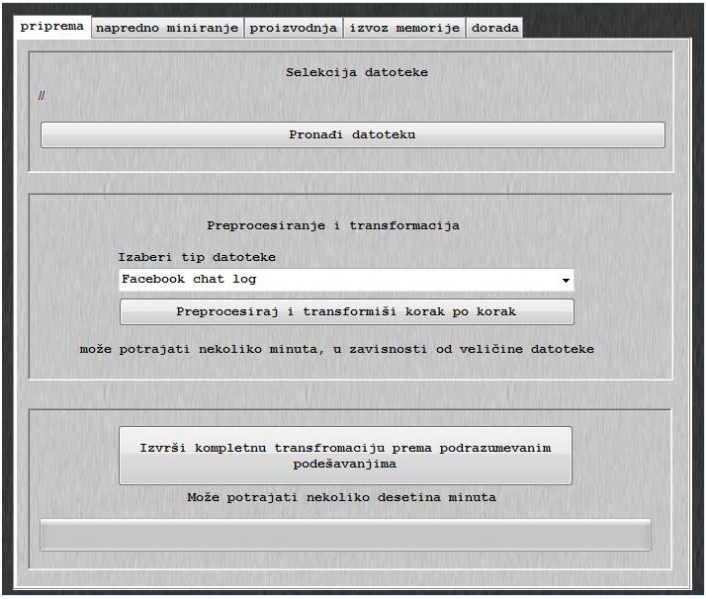
одређивач	значење	вредност	одређивач	значење	вредност
:))	смешак	0.56	:)	смешак	0.26
:D	осмех	0.91	:P	плажење	0.35
:p	плажење	0.24	xD	церење	0.59
:o	чуђење	-0.22	:O	чуђење	-0.29
:((	мрштење	-0.86	:(	мрштење	-0.66
:/	негодовање	-0.48	:\	негодовање	-0.48
:**	пољупци	0.88	:*	пољубац	0.80
hahaha	смејање	0.72	haha	смејање	0.15
<3	срце	1.00	;)	миг	0.26
:'(	плач	-0.74	lol	смејање	0.04
^^	радост	0.95	-.-	без речи	-0.45
:3	мачка	0.94	*.*	сјај у очима	0.95
:S	негодовање	-0.05	:'D	церекање	0.95
o.o	неверица	-0.10	...	без речи	-0.18

Табела 1. Коначна листа одређивача и њихових вредности.

3. Израда базе

За израду базе података задужен је био софтвер, направљен специјално за потребе овог истраживања. Написан је на програмском језику C# и може се покренути на *Windows* платформи, користећи било коју верзију овог оперативног система која ради на 64 бита. Корисничко окружење је предусретљиво и у потпуности на српском језику. Циљ при конципирању и изради овог софтвера био је да било који истраживач који говори српски може да га самостално користи, креира нове базе и употреби га за нова будућа истраживања.

Задатак софтвера је да улазну датотеку, која је у одговарајућем облику, кроз седам корака трансформише у базу података која садржи



Слика 2. Изглед корисничког окружења софтвера за истраживање података.

појмове који су у тој датотеци пронађени и њихове вредности на скали негативно-позитивно зависно од одређивача који се налазе у одговарајућој близини појма. У наредном поглављу биће објашњени кораци аутоматске израде базе података.

3.1 Припремна обрада

Припремна обрада обавља се над датотеком коју корисник сам бира притиском на дугме *Пронађи датотеку*. Тада се отвара *Windows Explorer* каталог у којем корисник на уобичајен начин бира датотеку са локалног рачунара. Алат тренутно подржава само датотеке које садрже историју дописивања корисника на друштвеној мрежи *Facebook*, и једини формат датотеке који корисник може да одабере је *htm*. Корисник наставља даље притиском на дугме *Препроцесирај и трансформиши корак по корак*[4] (слика 2).

[4] Алтернативно, може се користити дугме *Изврши комплетну трансформацију према подразумеваним подешавањима*, након чега

```
▼<html>
  ▼<head>
     <meta http-equiv="Content-Type" content="text/html; charset=UTF-8"/>
     <title>Korisnik 1 - Messages</title>
     <link rel="stylesheet" href="../html/style.css" type="text/css"/>
   </head>
  ▼<body>
   ▼<div class="nav">
      <img src="../photos/profile.jpg"/>
    ▶<ul>...</ul>
    </div>
   ▼<div class="contents">
      <h1>Korisnik 1</h1>
    ▼<div>
     ▼<div class="thread">
        Korisnik 1, Korisnik 2
      ▼<div class="message">
       ▼<div class="message_header">
          <span class="user">Korisnik 1</span>
          <span class="meta">Tuesday, February 9, 2016 at 2:44pm UTC+01</span>
        </div>
      </div>
      <p>Tekst poruke</p>
     ▼<div class="message">
       ▼<div class="message_header">
          <span class="user">Korisnik 2</span>
          <span class="meta">Tuesday, February 9, 2016 at 2:39pm UTC+01</span>
        </div>
      </div>
      <p>Tekst poruke</p>
```

Слика 3. Одломак улазне датотеке.

Како софтвер користи XSL трансформације, основни предуслов је да његов улаз буде добро формирана XML датотека. Датотека коју је корисник одабрао (слика 3) најпре се очисти од свих знакова који могу сметати при провери њене добре формираности. То се обавља коришћењем функције *Regex.Replace* и регуларног израза *[\u0000-\u001F]*, који проналази у датотеци појављивање прва 32 карактера *ASCII* скупа и замењује их празном ниском. Ради додатног обезбеђивања добре формираности нове датотеке карактер *&* се такође замењује размаком и уводи се нови корени елемент.

Припремна обрада се завршава трансформацијом структуре датотеке из почетне у жељену, која је обрадива (слика 4). Овом

ће се извршити комплетна трансформација према подразумеваним подешавањима или подешавањима коришћеним приликом последње трансформације корак по корак.

```
▼<root>
  ▼<conversation>
    ▼<messages>
      ▼<message>
          <sender>Korisnik 1</sender>
          <text>Tekst poruke</text>
        </message>
      ▼<message>
          <sender>Korisnik 2</sender>
          <text>Tekst poruke</text>
        </message>
```

Слика 4. Одломак датотеке након припремне обраде.

приликом се одстрањују сви чворови чија деца не садрже поруке који су корисници друштвене мреже разменили. Једини чвор који се чува је * * (слика 3) у којем су садржане поруке. Врши се преструктурирање ових чворова, нова датотека се чува и тиме се припремна обрада завршава.

3.2 Екстракција вредности из датотеке

Екстракција одређивача обавља се из два дела. Први део екстракције се састоји из задавања одређивача који ће се претражити и вредности које они означавају. Одређивачи и вредности се могу задати један по један, а овај корак се може и прескочити. Уколико се прескочи, софтвер ће аутоматски учитати подразумеване одређиваче и њихове вредности (табела 2.2).

У каталогу за напредно ископавање (text mining)(слика 5) изложени су сви подразумевани одређивачи и њихове вредности. Вредности се мењају ручно, путем текстуалног поља поред сваког од израза. Уколико је неки одређивач непожељан, његову вредност треба заменити карактером "/" и он неће бити коришћен приликом ископавања. У случају да корисник жели да тестира неки нови одређивач, може да користи дугме *додај емотикон* при дну странице, након што је попунио обавезна поља за израз и вредност. Уколико унета вредност одређивача није у опсегу од -1 до 1 програм је неће прихватити и корисник ће добити поруку о неуспешном додавању новог одређивача. Максимални број одређивача који се могу користити је 36.

Слика 5. Пример подешавања екстракције вредности из текста.

Напредни корисник може да измени текстуалну датотеку *data/emoticons.txt* која садржи подразумеване вредности одређивача. У сваком тренутку, одређивачи се могу вратити на подразумевана подешавања из поменуте датотеке, али се и нова подешавања могу преснимити преко постојећих. Обе опције се активирају притиском на одговарајуће дугмиће који се налазе на дну екрана (слика 5). У оба случаја, због могућег неповратног губитка података, корисник ће морати да потврди процес у додатном дијалогу који ће се појавити на екрану.

Након што је корисник задовољан подешавањима може прећи на други део екстракције који се обавља притиском на дугме *екстрактуј вредности из текста* на картици за производњу (слика 7). Пре притиска на ово дугме корисник може, али то није обавезно, да одабере поље *Користи предодређене регуларне изразе за побољшану претрагу*, што би требало да доведе до екстракције већег броја одређивача из текста.

Извршавање започиње провером обележености опционог поља. Уколико је обележено, текст пролази кроз низ *Regex.Replace* функција које проналазе унапред позната одступања од уобичајеног исписивања неколицине одређивача. Најпре се исправљају аутоматска превођења карактера у XML-у. Како овај део трансформације не захтева добро формиран XML документ, ентитетска референца *lt;* (ознака за почетак ентитетске референце *&* је приликом препроцесирања замењена размаком) замењује се карактером *<*, како би емотикон *<3* могао бити пронађен у тексту. Ентитетска референца *039* замењује се карактером *'*, како би емотикони попут *:'(* могли бити пронађени. Ниска *:-* замењује се самосталним карактером *:* , како би у ископавање били укључени и емотикони са *носем* типа *:-D* или *:-(*. Сва ћирилична слова која се користе у унапред одабраним одређивачима се транслитеришу у латиницу како би се ти емотикони пронашли и код корисника који користе ћирилицу (*:Д*, *хД*). Све варијанте емотикона *xD*, *:S* и *o.o* записане великим и малим словима, претварају се у основни формат. Почеци хиперлинкова *http://* и *https://* замењују су неутралном карактерском ниском *yy*, како не би били погрешно протумачени као емотикон *:/* . Коначно, регуларни изрази *[a|h|A|H][a|h|A|H][h|H][a|A][h|H][a|h|A|H]* и *[h|H][a|A][h|H][a|A]* се користе како би било пронађено што више различитих примера изражавања смеха, а пронађени изрази се замењују са *hahaha* односно *haha*.

Уколико опција чишћења и уједначавања није укључена, овај корак се прескаче и биће пронађени и обрађени само одређивачи у свом основном облику који је наведен на картици напредног ископавања.

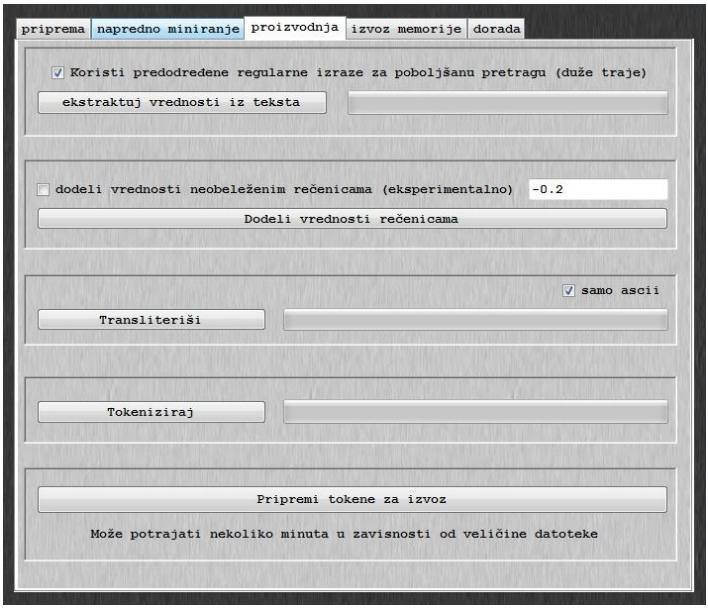
Сама екстракција се састоји од учитавања одабраних израза и њихових вредности са картице напредног ископавања у два низа и потом њихове обраде. Сваки од елемената првог низа, низа израза, пролази кроз *for each* петљу у којој се свако његово понављање у тексту замењује етикетом *<emot value='x'/>*, која представља један празан XML чвор у коме је x вредност текућег одређивача која је преузета из другог низа, низа вредности. Уколико је вредност израза */* инструкције неће бити извршене, а уколико вредност неког израза није између -1 и 1 програм ће пријавити грешку и прекинути извршавање.

Након што је сваки од одређивача замењен чвором чији је атрибут његова вредност (слика 6), XML документ поново постаје добро формиран и спреман за даљу обраду. Корисник ће добити поруку да је екстракција завршена и да може да пређе на следећи корак.

```
▼<message>
   <sender>Korisnik 1</sender>
   <text>tekst :)</text>
 </message>

▼<message>
   <sender>Korisnik 1</sender>
   <text>teskt <emot value="0.26"></text>
 </message>
```

Слика 6. Изглед поруке пре и након екстракције одређивача.



Слика 7. Картица за производњу базе података корак по корак.

3.3 Додела вредности сегментима текста

У експерименту сегмент текста или реченица поистовећена је са једном самосталном поруком коју је један корисник упутио другом.

Основна идеја је да се вредности одређивача пронађених у поруци одразе на вредност саме поруке, или околних порука у специјалним случајевима. У зависности од текстуалног садржаја поруке (осим самог одређивача[5]) и садржаја претходне и наредне поруке постоји три врсте доделе вредности сегментима:

1. уколико порука поред одређивача садржи текст и порука која је следи садржи текст – на поруку се односе само одређивачи из те поруке. Примери:
   *А: Срећан рођендан :**
   Б: Хвала пуно
   На поруку особе А односиће се одређивач *:**, док порука особе Б на њу нема утицај јер не садржи одређивач. *А: Срећан рођендан :**
   Б: Хвала :)
   На поруку особе А односиће се одређивач *:**, док порука особе Б на њу нема утицај јер се одређивач у поруци особе Б односи само на текст поруке особе Б.

2. уколико порука садржи тескт али не и одређивач, а наредна порука садржи одређивач, али не и текст – одређивач ће се односити на тескт поруке која претходи. Пример:
   А: Аутобус ми је побегао
   Б: :(
   Одређивач *:(* из поруке особе Б односиће се на поруку особе А јер порука особе Б не садржи текст, те се њен одређивач односи на поруку особе А, која јој претходи у разговору.

3. уколико порука садржи и одређивач и текст, а наредна порука садржи одређивач, али не и текст – одређивачи из обе поруке се односе на поруку која садржи текст. Пример:
   А: Аутобус ми је побегао -.-
   Б: :(
   На поруку особе А односиће се и одређивач -.- и одређивач *:(*, јер порука особе Б не садржи текст, те се њен одређивач односи на поруку особе А, која претходи у разговору.

[5] У овом случају све поруке које садрже мање од 4 карактера третирају се као празне.

Постоји и четврта могућност – додела вредности поруци која не садржи одређивач, при чему ни следећа порука не садржи одређивач. Ова могућност се заснива на претпоставци да и одсуство одређивача само по себи носи неко значење (према интуицији негативно). Недостатак ове могућности је да не можемо бити сигурни да непостојање одређивача носи (негативно) значење, док је одређивање саме вредности тежак задатак. Додатни проблем је и то да у поруци можда и постоји одређивач који кориснику није познат па га није користио при ископавању тескта. Стога је ова опција необавезна и њена вредност није дефинисана. Уколико корисник жели да зада одређену вредност *необележеним* порукама, то може да уради одабиром поља *додели вредност необележеним реченицама* и исписивањем вредности између -1 и 1 у текстуално поље пре покретања саме доделе (слика 7).

```
▼<root>
  ▼<conversation>
    ▼<messages>
      ▼<message>
          <sender>Korisnik 1</sender>
          <text>tekst poruke 1 <emot value="0.26"><emot value="-0.26"></text>
        </message>
      ▼<message>
          <sender>Korisnik 2</sender>
          <text><emot value="0.15"></text>
        </message>
```

```
▼<root>
   <emotext value="0.15">tekst poruke 1</emotext>
```

Слика 8. Изглед одломка документа пре и после доделе вредности сегментима када и наредна порука садржи одређивач, али не и текст (случај 3).

Сама додела вредности извршава се путем XSLT трансформације. Најпре се за сваку *обележену* поруку рачуна аритметичка средина вредности свих одређивача коју се на њу односе, затим се та вредност уписује у њен чвор у нови атрибут *value*, а чворови који су обележавали одређиваче се бришу. Потом, уколико је опционо поље било обележено, осталим порукама се додаје задата вредност. Резултат трансформације

је XML документ који садржи корени елемент и у њему чворове свих порука које имају додељену вредност. По завршетку трансформације корисник ће добити поруку да је додела завршена и зелено светло за наставак даље.

3.4 Транслитерација и сравњивање

Ова два корака обављају се заједно притиском на дугме *транслитериши* картице за производњу. Сврха овог дела програма је да се добије чист текст за прављење базе. У овом случају текстом се сматрају ниске које садрже искључиво бројеве и латинична слова. Необавезно поље за избор *само ascii*, служи да се скуп додатно ограничи само на латинична слова из ASCII скупа карактера (бројеви се овом опцијом не изузимају).

Разлог због којег је превођење у ASCII необавезно је што се њиме уводи готово исто толико проблема колико се и решава. Под претпоставком да сви корисници друштвених мрежа не користе у пуној мери капацитет Unicode каракатерског скупа, већ и *деградирану латиницу*,[6] долази до проблема јер неке речи попут *истраживање* и *истразивање* неће бити препознате као иста, већ као две различите речи. Превођењем свих карактера у ASCII скуп овај проблем се избегава али се уводи нови проблем да речи које би се заправо разликовале у Unicode скупу буду препознате као иста реч – пример могу да буду речи *шпанац* и *спанаћ*, које ће програм препознати као исте речи. Ову опцију је, дакле, најбоље користити зависно од ситуације.

Пре даљег наставка обраде, тренутни XML документ пролази кроз XSL трансформацију у којој се помоћу функције *translate* сва велика слова претварају у мала, како програм не би у коначној бази правио разлику између две исте речи, које се разликују једино у величини слова.

Други део сређивања текста, чишћење свих карактера који нису слова или цифре, отклања могућност да реч садржи неки интерпукцијски или други неалфанумерички карактер. Поново се функцијом *translate* XSL трансформације сви нежељени карактери (сви карактери осим малих латиничних слова и цифара) претварају у размак. И овим поступком могу се добити некоректне речи ако је корисник случајно унео непожељни карактер у току куцања речи (*Мар?ија*) или

[6] Латинична слова без дијакритичких знакова, на пример: *c*, *z*, *s* уместо *č*, *ž*, *š*.

ако је намерно користио неки специјални карактер (*M@puja*). У првом случају ниска ће постати *Map uja*, а у другом *M puja*.

3.5 Токенизација

Токенизација текста обавља се притиском на дугме *токенизација* картице за производ (слика 7). Подела на токене се врши искључиво према размаку, па се токен дефинише као било која ниска карактера између почетка поруке и првог размака у поруци, било која ниска карактера између два суседна размака или било која ниска карактера између последњег разамака у поруци и краја поруке. Она се у овом случају врши на нивоу XML документа, тако што сваки токен добија сопствени XML чвор. Први део процеса токенизације обавља се функцијом *Regex.Replace* у три корака:

- Најпре се проналази сваки почетак поруке у XML документу. То се обавља претрагом ниске *">*, која се јавља једино на крају отворене етикете елемента који садржи атрибут, а то је елемент који садржи текстуалну поруку. Како корени елемент не садржи атрибут, он неће бити захваћен овом претрагом. Затим се после сваке пронађене ниске додаје ниска *<token>*, која означава почетак токена.
- Потом се проналази крај сваке поруке, претрагом ниске *</emotext>* што је етикета затварања елемента поруке. Испред сваке овакве пронађене ниске додаје се *</token>*, како би се добила ниска *</token></emotext>*. Овим кораком свака порука у XML документу постаје један токен: *<emotext value="x"><token> ... </token></emotext>* .
- Последњи корак је подела унутрашњости сваке поруке на токене, на сваком месту где се налази размак. Како не би дошло до нежељене замене унутар отворене етикете *<emotext value="x">*, на месту између назива елемента и назива атрибута, најпре се обавља замена сваког *emotext value* са *emotext_value*, а тек онда замена сваког размака са *</token><token>*. Након тога се карактер _ поново враћа у размак и добијају се поруке у облику: *<emotext value="x"><token> ... </token><token> ... </token></emotext>* .

Други део токенизације обавља се помоћу четири додатне XSL трансформације које филтрирају одабране токене и дају им вредност:

– Прва трансформација брише све поруке које садрже више од четрдесет токена, како би се избегла додела вредности свим речима у, на пример, неком тексту налепљеном уз ћаскање, а који није активна порука у конверзацији.
– Друга трансформација брише све токене који садрже мање од два карактера (под претпоставком да не носе значење јер представљају нефункционалне речи или су настали разбијањем великих ниски попут линкова), и све токене који садрже више од двадесет карактера (под претпоставком да је велика већина њих насумична).

Између ове две трансформације потребно је спровести још један корак, увођење додатног токена без вредности на последње место у документу. То се ради уметањем ниске *<token>ERR0001</token>* пре затворене етикете кореног елемента, како би се добило *<token>ERR0001</token>*. Значење ове ниске биће објашњено у наредном одељку.

– Трећа трансформација служи за осамостаљивање токена. Она најпре сваком токену додељује атрибут са вредношћу поруке у којој се налази, а затим брише елементе реченица, остављајући у кореном чвору само чворове токена са придруженим вредностима. Она такође сортитра све токене по абецедном реду у односу на њихов текстуални садржај и додељује сваком токену нови атрибут *no*, чија вредност постаје редни број токена у документу, одређен XSL функцијом *position*.

```
▼<root>
   <emotext value="0.15">tekst poruke 1</emotext>
```

```
▼<root>
   <tok val="0.15" no="23656">poruke</tok>
   <tok val="0.15" no="28649">tekst</tok>
```

Слика 9. Изглед одломка документа пре и након токенизације.

– Четврта трансформација не доприноси садржају документа већ само економичности његове обраде. Име сваког елемента *token* се скраћује на *t*, сваког атрибута *value* на *v*, што у великој мери смањује величину документа и олакшава и убрзава даљи рад са датотеком.

3.6 Израда индекса токена са вредностима

Овај корак се обавља притиском на дугме *припреми токене за извоз* у задњем пољу картице за производњу (слика 7). Извршење ове команде могуће је тек након што су извршени сви претходни кораци припреме.

Како би индекс послужио сврси, сви појмови у њему добијају два атрибута: средња вредност свих одређивача који се на тај појам односе у учитаном корпусу и број понављања тог појма у њему. Креирање индекса обавља се у шест корака XSL трансформацијом XML документа који у себи садржи токене:

- У бази уређеној у абецедном поретку се проналази први чвор, x, и први следећи чвор који се од њега разликује, y.
- Узимају се њихови редни бројеви токена из атрибута *po*. Редни број чвора x је n, а редни број чвора y је m.
- Текст чвора x се уписује у нови XML документ инвертованог индекса у елемент t (token).
- Новом елементу t додељује се атрибут c (count), који означава број понављања тог токена у бази, а његова вредност је једнака разлици редних бројева два пронађена чвора.

$$c = m - n \tag{2}$$

- Новом елементу t додељује се атрибут v (value), који означава вредност тог токена на скали позитивно-негативно, и она је једнака аритметичкој средини вредности из атрибута v свих токена који садрже исти текст.

$$v_x = \frac{\sum_{i=n}^{n+c-1} v_i}{c_x} \tag{3}$$

- Чвор y постаје први чвор у документу и процедура креће из почетка.

Овим поступком атрибути се додељују свим токенима, сем последњем, помоћном, *ERR0001* токену, који је створен приликом токенизације документа и служи за обележавање завршетка документа и за израчунавање атрибута c код токена који се налази испред њега.

Након што је успешно извршено претварање скупа токена са вредностима у коначан инвертовани индекс појмова са вредностима (табела 3.6), корисник ће добити поруку о успешно извршеној трансформацији, а индекс ће бити сачуван у привременој датотеци док се не изврши његов коначан извоз.

t (појам)	v (вредност)	no (редни број)
zdravko	-0.1	859231
zdravko	0.3	859232
zdravlje	0.26	859233

t (појам)	v (средња вредност)	c (број понављања)
zdravko	0.2	2
zdravlje	0.26	1

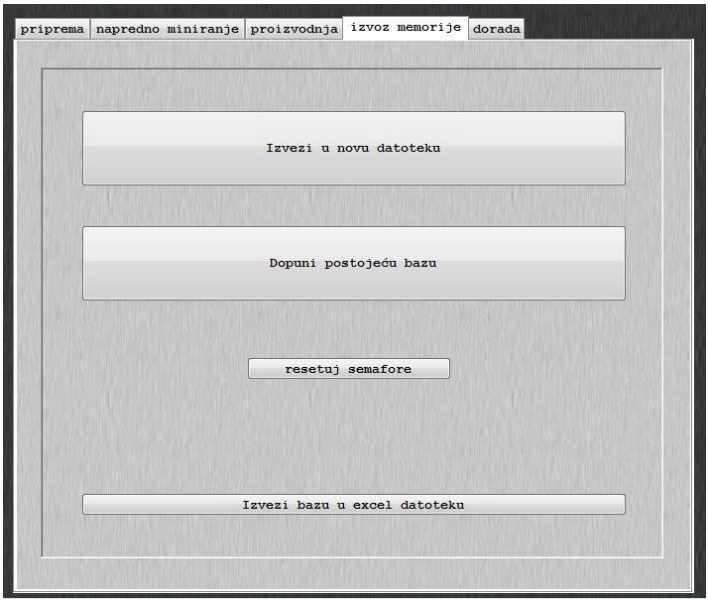
Табела 2. Пример изгледа дела базе података пре и након коначне трансформације.

4. Извоз и допуна базе података

Притисак на дугме *Извези у нову датотеку* картице за извоз меморије (слика 10) отвара *Windows Explorer* каталог за чување нове датотеке. Формат у коме ће датотека бити сачувана је XML, а корисник бира њено име и место. Садржај се копира из привремене датотеке креиране у претходном кораку, и представља коначну базу података.

У случају да је корисник већ обрађивао неке датотеке, на располагању му је и опција *Допуни постојећу базу* (слика 10). Она омогућава да се подацима извученим из новог корпуса допуни претходно извезена база података. У овом случају, притиском на одговарајуће дугме отвара се *Windows Explorer* каталог за отварање нове датотеке. Корисник треба да обележи датотеку коју жели да допуни, а затим се креира допуњена база у четири корака:

– Изабрана датотека се у потпуности учитава у меморију и додаје на садржај претходно креираног индекса сачуваног у привременој датотеци.
– Проналази се и брише ниска *</root><root>*. Тиме се добија документ са једним, уместо два корена елемента.
– XSL трансформацијом се све токени поново сортирају по абецедном реду.
– Како у овом случају могу постојати највише два токена са истим текстом, њихова унификација је једноставнија. Уколико токен нема парњака, преписује се, а уколико има добија само нове вредности атрибута. Атрибут вредности се израчунава као количник збира производа атрибута v и c оба елемента и укупног броја њихових



Слика 10. Изглед картице за извоз/допуну базе података.

понављања (2.4). Атрибут понављања једнак је укупном броју њихових понављања (2.5).

$$v = \frac{v_1 * c_1 + v_2 * c_2}{c_1 + c_2} \tag{4}$$

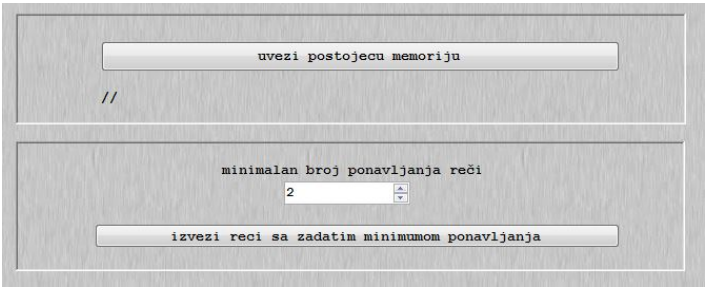
$$c = c_1 + c_2 \tag{5}$$

Нови елемент са атрибутима се чува док се стари бришу из индекса, и тиме се добија по једна копија сваког токена са правилно одређеним вредностима атрибута.

Нови документ замењује онај који је обележен у каталогу за отварање, те је ово опција коју треба употребљавати пажљиво. По завршетку, корисник ће добити поруку о успешно извршеном допуњавању.

Уколико корисник жели да ручно прегледа новостворену базу, а не осећа се пријатно у XML окружењу, може да је извезе у CSV (engl.

comma separated values – вредности раздвојене зарезом) формат који може да се отвори помоћу *Microsoft Excel*-а, *Open Office Calc*-а или сличног софтвера за рад са табелама. Он то чини притиском на последње четврто дугме на картици за извоз меморије.

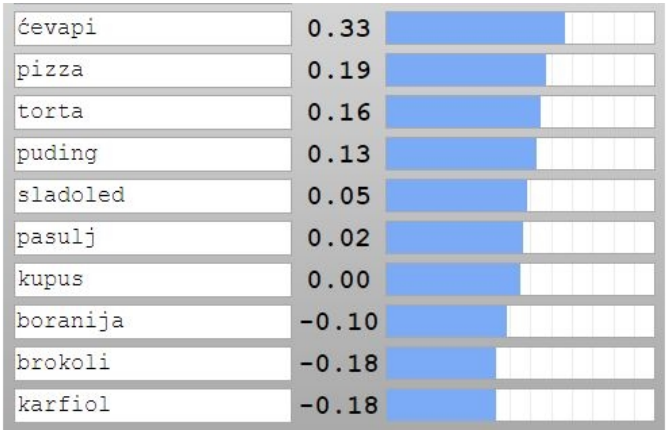


Слика 11. Изглед одељка за дораду базе података

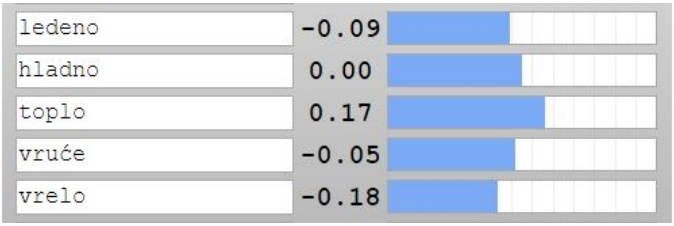
База података се може додатно проредити на основу тога колико често се појмови понављају. Ова опција се налази на петој картици софтвера, експерименталној картици за дораду. Корисник најпре треба да учита базу података коју жели да доради притиском на дугме *увези постојећу меморију*, а затим бира колико барем пута појам треба да се понавља да би био узет у обзир (слика 11).

Уколико корисник, на пример, одабере број 2, притиском на дугме *извези речи са задатим минимумом понављања* формираће се нова база из које ће бити избрисани сви појмови који се понављају мање од 2 пута (тј. појављују се само једном), и нова база ће бити сачувана на жељеном месту. Ову опцију ће корисник користити уколико сматра да је потребно да се појам појави одређени број пута како би његова вредност могла да се сматра репрезентативном.

Након што је корисник завршио креирање жељених база и успешно их извезао, може их тестирати ручно уз помоћ претраге уграђене у било који софтвер за рад са текстом или XML документима. Такође, база се у било којем од својих облика може тестирати и у софтверу за проверу вредности прављеним специјално за овај експеримент (слике 12 и 13).



Слика 12. Пример поређења вредности појмова из базе података коришћењем веб-апликације – брза храна, дезерти, поврће.



Слика 13. Пример поређења вредности појмова из базе података коришћењем веб-апликације – температуре.

5. Закључак

5.1 Осврт на резултат експеримента

Упркос релативно малом коришћеном узорку од 1.843.826 порука, од грубо процењених 500 милијарди порука које припадају корисницима друштвене мреже *Facebook* у Србији, експеримент је испунио своје циљеве. На основу екстрахованих свега двадесет и осам одређивачких ниски уз помоћ седам регуларних израза који проналазе њихове варијације, и без претходног знања о језику и његовој граматици,

софтвер је формирао базе података са појмовима и њиховим вредностима на скали позитивно-негативно.

Систем је тестирало неколико независних оцењивача и на основу њихових реакција може се закључити да су вредности ненасумичне и репрезентативне, што показује да је овај начин класификовања појмова могућ у било којем природном језику који користи одређиваче, али је и даље неопходно спровести системску евалуацију.

Коришћени алгоритми су и даље далеко од савршених и у њима има доста простора за побољшавање и напредак. Добра документованост процеса истраживања података требало би да допринесе будућим истраживањима исте или сличних идеја. Следећи корак свакако може да буде екстраховање проширеног скупа одређивача, рад над проширеним скупом текстова или, идеално, проширени скуп параметара који би се вредновали. Корпус проширен, на пример, на друге природне језике умногоме би допринео обиму појмова којима се вредност додељује, док би проширени скуп параметара могао да дода нову дубину разумевања текста од стране како машине тако и људи.

5.2 Могућност примене

Овај начин екстракције одређивача и уопште сличног истраживања података може наћи разноврсне примене које се могу поделити у две групе.

Друштвена и демографска истраживања:

- Маркетиншка истаживања: истаживање деловања тренутног или алтернативног приступа рекламирању производа и услуга. Ово је најчешћа употреба сличних истраживања пре свега из новчаних разлога, јер компанијама може да уштеди новац или да им обезбеди нови прилив добара.
- Истраживање јавног мњења: шта људи воле, шта не воле, какво уопште мишљење имају о стварима или идејама на које се појмови или текст односе. Може се употребити на различите начине у друштвеним и демографским истраживањама за брзо и ефикасно прикупљање велике количине информација.

Развијање интелигентних система за рад са информацијама:

- Проналажење информација: проналажење специфичних информација у тексту, као и проналажење информација које

се не могу прецизно дефинисати. Класификација текстова према расположењу које је у њима изражено ради лакшег проналажења потребних информација.
- Анализа и разумевање природног језика: разумевање писаног текста и текстуалних упита, анализа расположења које текст носи, рад са дигиталним језичким ресурсима попут аутоматизоване паралелизације или аутоматизација било које операције која захтева дубоко разумевање писаног текста.
- Вештачка интелигенција: аутоматизовање конверзације на природном језику, одржавање разговора и рад са странкама.

5.3 Могући недостаци и побољшања

- Проблем могућих насумичних одговора приликом анкетирања:
  Направити систем који одбацује одговоре корисника који су одговарали насумично, на пример, увођењем додатних питања која у тексту захтевају специфичне одговоре. Тако немарни корисници који не читају питања имају веће шансе да буду идентификовани.
- Системска евалуација резултата:
  Спровести детаљну системску евалуацију како би се утврдила веродостојност добијених резултата.

Литература

Davidov, Dmitry, Oren Tsur и Ari Rappoport. “Enhanced Sentiment Learning Using Twitter Hashtags and Smileys”. У *COLING ’10 Proceedings of the 23rd International Conference on Computational Linguistics: Posters*, Association for Computational Linguistics, 2010, 241–249.

Go, Alec, Richa Bhayani и Lei Huang. “Twitter Sentiment Classification Using Distant Supervision”. *CS224N Project Report, Stanford* Vol. 1 (2009).

Mishne, Gilad. “Experiments with mood classification in blog posts”. У *Proceedings of ACM SIGIR 2005 workshop on stylistic analysis of text for information access*, Vol. 19, Citeseer, 2005, 321–327.

Neviarouskaya, Alena, Helmut Prendinger и Mitsuru Ishizuka. “Compositionality Principle in Recognition of Fine-Grained Emotions from Text”. У *Proceedings of the Third International ICWSM Conferenc*. The AAAI Press, 2009, 278–281.

Ptaszynski, Michael, Pawel Dybala, Rafal Rzepka и Kenji Araki. “Towards Fully Automatic Emoticon Analysis System”. У *Proceedings of the Sixteenth Annual Meeting of the Association for Natural Language Processing*, 2010, 223–228.

Ptaszynski, Michael, Rafal Rzepka, Kenji Araki и Yoshio Momouchi. “Research on Emoticons: Review of the Field and Proposal of Research Framework”. У *The Seventeenth Annual Meeting of The Association for Natural Language Processing*. The Association for Natural Language Processing, 2011, 1159–1162.

Нормативна контрола у Србији

УДК 025.342:006.44(497.11)
УДК 025.342:004.65(497.11)

Ана Савић
ana.savic@nb.rs
Народна библиотека Србије

САЖЕТАК: Рад описује процес стварања српске нормативне датотеке личних имена, њеног значаја за каталогизацију, развој универзалне библиографске контроле и српске библиотечке заједнице. Разматра се нормативна контрола у 21. веку у свету и описује се као оруђе помоћу којег библиотеке учествују у агенди повезаних отворених података.

КЉУЧНЕ РЕЧИ: нормативна контрола, српска нормативна датотека личних имена, повезани отворени подаци

РАД ПРИМЉЕН: 13. септембар 2016.
РАД ПРИХВАЋЕН: 15. децембар 2016.

1. Увод

У својој вековној традицији, библиотеке и њима сродне установе првенствено су биле оријентисане на сакупљање и каталошку обраду резултата интелектуалног или уметничког рада особа и институција. Једнозначно одређивање особа и институција повезаних са делом које се описује, као и опис публикације, били су кључни приликом каталошке обраде ради проналажења и приступања описаном извору (Angjeli et al., 2014). Идентификација особа повезаних са изворима у каталозима произилазила је из усредсређености на избор имена и његовог облика које је основа јединствене одреднице, формулисања јединствене одреднице и тежње да се успостави синдетичка структура каталога,[1] то јест, да се успостави веза између заједничких карактеристика окупљањем библиографских јединица повезаних истим аутором или

[1] Синдетичка структура каталога подразумева успостављене везе између система „види“ и „види и“ упутница и одредница, то јест, упућује на каталоге са развијеним везивним механизмима.

особом којој се приписује неко ауторство. Са концептом повезаних података (linked data) и семантичког веба модел top-down, у који су се библиографски записи електронских каталога на MARC формату уклапали и одговарали потребама тадашњих корисника и окружења деведесетих година 20. века, више није био довољан за опстанак библиотечких каталога који су морали постати део стварности веба 2.0. „Потреба да се побољша интероперабилност глобалне мреже условила је развој семантичког веба који је, са друге стране, довео до стварања бројних нових начина контроле и стандардизације описа докумената, до решавања проблема који су оптерећивали различите системе за индексирање, као и до побољшања интероперабилности записа (SKOS-а, SIOC-а, Даблинског језгра, FOAF-а и других)“ (Леива-Медерос и др, 2015). Један од начина преобликовања каталога у том смислу је приљежан рад на стварању и развијању нормативних датотека и њихово повезивање са библиографским датотекама. Пројекти нормативних датотека почели су да се развијају под окриљем универзалне библиографске контроле почетком седамдесетих година 20. века када је усвојена препорука да „свака национална библиографска агенција треба да одржава систем нормативне контроле за имена аутора, лична и колективна, и за јединствене наслове у складу са међународним смерницама“ (Guidelines, 1979). У складу са тим, једнозначна идентификација особа повезаних са извором који се описује на националном нивоу, то јест, националне нормативне датотеке личних имена за новонастало окружење семантичког веба постало је изузетно вредно оруђе повезивања и размене информација у првом кораку стварања међународних датотека (на пример, VIAF[2] и ISNI[3]), а затим и поуздан извор за употребу која превазилази оквире библиотечке заједнице.

2. Увођење нормативне контроле у библиотечко-информациони систем COBISS

Кооперативни онлајн библиографски систем и сервиси COBISS развијају се од 1987. године када је успостављен систем узајамне

[2] VIAF (Virtual International Authority File) – Виртуелна међународна нормативна датотека, https://viaf.org/

[3] ISNI (International Standard Name Identifier) – Међународна стандардна именска ознака, http://www.isni.org/

каталогизације на нивоу целе некадашње Југославије. Носилац развоја организационих решења и програмске опреме био је Институт информационих знаности (ИЗУМ) из Марибора. Као надоградњу постојећег система узајамне каталогизације, ИЗУМ је 1991. године промовисао COBISS. Услед друштвених и политичких промена COBISS је наставио да се развија као словеначки библиотечко-информациони систем, а када је на библиотечком плану поново успостављена сарадња са новонасталим државама на територији бивше Југославије, 2003. године постигнут је споразум о успостављању мреже COBISS.Net што је значило повезивање националних база података пет држава (Словенија, Србија, Босна и Херцеговина, Македонија и Црна Гора).

Још током развоја система као националног словеначког система, при чему је у средишту пажње било повезивање локалних база у узајамну COBIB базу, конзистентност, уједначеност и квалитет како локалних база тако и узајамне базе одржавали су се „нормативном контролом личних имена аутора, контролом дуплираних записа, програмским контролама за формате COMARC, редакцијом записа, глобалним шифрарницима за све стандардизоване податке (нпр. државе, језици, УДК), локалним шифрарницима за све податке који су уједначени унутар библиотеке (нпр. локације, интернеознаке), аутоматским нумераторима (нпр. инвентарски бројеви, бројеви у сигнатури), контролом једнозначне идентификације серијских публикација итд.“[4]. Међутим, пројекат нормативне контроле није се развијао упоредо са библиотечко-информационим системом.

Како наводе М. Сељак и др., прве анализе спроведене су 1994. године, а прва верзија прототипа нормативне базе направљена је 1996. године када је узајамна библиографска база садржала око милион библиографских записа. Оцењујући квалитет нормативне базе као незадовољавајући, ИЗУМ је наставио да надграђује прототип нормативне базе па је направио три верзије. CONOR V3.0 је промовисан 1998. године и представљао је солидну основу за даљи развој пројекта нормативне контроле. Уследила су разматрања редакције записа, организације рада, припремање приручника за нормативне податке, предлози начина повезивања библиографских и нормативних записа и предлози за надоградњу програмске опреме. У сарадњи са Националном

[4] Народна библиотека Србије, „Platforma *COBISS – Kooperativni online bibliografski sistem i servisi* “http://www.cobiss.net/platforma_cobiss-SR.htm (преузето 22. 07. 2016.)

и универзитетском библиотеком (Nacionalna in univerzitetna knjižnica – NUK) у Љубљани 2000. године почело је тестирање четврте верзије нормативне базе CONOR. Тек је верзија из следеће године почела да се редигује и допуњује новим записима и 2003. године укључена је у систем узајамне каталогизације Словеније. Та верзија се темељила на избору одређеног броја најквалитетнијих библиографских записа који су били основ за програмско креирање нормативних записа. Најпре су креирани нормативни записи личних имена словеначких аутора изабраних из корпуса издавачке продукције последњих неколико година на основу персоналних библиографија и словеначке националне библиографије монографских публикација. План је био да се сви програмски креирани нормативни записи редигују пре укључивања нормативне базе у систем узајамне каталогизације. Увођење нормативне контроле подразумевало је и надоградњу програмске опреме за узајамну каталогизацију. У јануару 2003. године развијена је верзија програмске опреме COBISS2/Katalogizacija V6.0 за узајамну каталогизацију са нормативном контролом и започета је обука каталогизатора. Настављено је тестирање програмске опреме и приређено је ново издање приручника COMARC/A на основу издања из 1998. године. Планирани развој нормативне контроле односио се на увођење нормативне контроле за називе колективних тела тако да данас нормативна база личних имена аутора и колективних тела Словеније садржи преко 900.000 нормативних записа(Seljak, 2004).[5]

Увођење нормативне контроле у систем узајамне каталогизације Словеније био је значајан помак у повезивању и омогућавању приступа подацима, обогаћивању квалитета националног библиотечко-информационог система и учвршћивања партнерског односа са сродним међународним системима. Пројекат је показао да је увођење нормативне контроле у библиографске базе података које су првобитно креиране без ње изузетно сложен посао и да је оптимално било да се пројекти узајамне каталогизације и нормативне контроле развијају упоредо. Такав концепт је економичнији, захтева мање времена и велики је професионални изазов за каталогизаторе.

[5] http://cobiss4.izum.si/scripts/cobiss?ukaz=BASP&bno=509&id=1955089254138144

3. Нормативна контрола у Србији

На основу предлога пројекта увођења нормативне контроле у систем COBISS.SR Љиљане Ковачевић и Гордане Поповић-Мазић у јануару 2006. године, а на иницијативу Центра Виртуелне библиотеке Србије (ВБС), израђен је документ *Uputstva za pripremu bibliografskih zapisa za programsko kreiranje normativnih zapisa za lična imena.* То је означило коначну одлуку да надоградња система узајамне каталогизације под платформом COBISS иде ка увођењу нормативне контроле. Природно, по усвајању предлога, реализација пројекта је започета у сарадњи са вишедеценијским партнером ИЗУМ-ом, а носиоци пројекта били су Народна библиотека Србије, Библиотека Матице српске и Универзитетска библиотека „Светозар Марковић“, уз партиципирање свих библиотека у Србији које одржавају и развијају своје каталоге под платформом COBISS. Ваља споменути да COBISS не окупља све библиотеке српске библиотечке мреже и увођење нормативне контроле не односи се директно и апсолутно на све локалне каталоге, али је стратешки и технолошки ово решење било оптимално. COBISS је једини имао потенцијал да буде база изградње и ширења савремене националне нормативне контроле у Србији за коју се претпостављало да ће најбрже дати резултате.

Акценат је стављен на именске одреднице с обзиром на то да је одлучено да се најпре креира нормативна датотека личних имена, уз даље проширивање на нормирање свих осталих приступних тачака. Крајем 2008. године Редакторски тим ВБС-а за монографске публикације разматрао је спорно питање такозваних „паралелних одредница“ и донео одлуку о примени овог решења у српској нормативној бази. Тада и почиње прва фаза реализације пројекта увођења нормативне контроле у Србији и она је подразумевала припрему и обележавање библиографских записа, и сачињавање корпуса записа који ће служити као основа за екстраховање нормативних записа. На реализацији ове фазе пројекта ангажовани су каталогизатори из Народне библиотеке Србије, Библиотеке Матице српске и Универзитетске библиотеке „Светозар Марковић“ и формирани редакторски тимови за вођење ових послова. Успостављање „паралелних одредница“ је спецификум српске нормативне датотеке личних имена и односи се на паралелну примену ћириличког и латиничког писма при формирању јединствене одреднице за лично име. У сличној ситуацији су и националне библиотеке других држава, на пример, Национална

библиотека и архив Канаде која има такозвану билингвалну каталошку политику. За особе чија се имена у етимолошком облику појављују и на енглеском и на француском језику креирају се „паралелне одреднице“.

Пажња је усмерена на примарно ауторство и коауторство, то јест, на податке унете у поља 700 и 701 COMARC/B формата. Предност је дата домаћим ауторима и страним ауторима чија су дела преведена на српски језик. У изабраним библиографским записима требало је прецизно и у складу са важећим каталошким правилима[6] утврдити јединствену одредницу аутора и све облике његовог имена. Угледајући се на словеначки модел, у Србији се тежило развијању сопственог модела што је било неопходно због поменуте специфичности која се огледа у паралелним облицима јединствених одредница. Наиме, наслеђе лисних алфабетских каталога библиотека у Србији је рефлектовано на узајамни електронски каталог COBIB.SR и употребу ћириличног и латиничног писма при избору и формирању јединствене одреднице личног имена. Стога је било неопходно успоставити принцип „паралелних одредница“ што је подразумевало ћирилични и латинични облик имена, али и упаривање фонетског и етимолошког облика имена за стране ауторе.

Пример 1

```
200 1<7>cb - ћирилица - српска<a> Младићевић <b> Жељко <f> 1977-
200 1<7>ba - латиница  <a>Mladićević  <b> Željko<f> 1977-
```

Пример 2

```
200 1 <7>ca - ћирилица - није специфицирана <a> Алексеев <b>
Глеб Васильевич <f> 1892-1938

200 1 <7>ba - латиница <a>Alekseev<b>GlebVasil’evič <f> 1892-1938

400 1 <7>cb - ћирилица - српска <9>srp - српски <a>Алексејев <b>
Гљеб Васиљевич <f> 1892-1938
```

[6] *Pravilnik i priručnik za izradbu abecednih kataloga. Dio 1, 2* у Србији је и даље важећи правилник за каталогизацију, тј. прихваћен је као национални каталошки правилник. Уз извесне измене које се односе на правила за формирање јединствених одредница личног имена аутора чији је етимолошки облик на страној ћирилици, нормирани облици имена формирају се у складу са првим томом поменутог Правилника.

Пример 3

```
200 1 <7>cb - ћирилица - српска <a> Томас <b> Скарлет <f> 1972-
200 1 <7>ba - латиница<a> Thomas <b> Scarlett <f> 1972-
```

Пример 1 илуструје „паралелну одредницу“ формирану за српског аутора која се састоји од ћириличног и латиничног облика имена. Када је реч о ауторима који стварају на језику који није српски и етимолошки, облик њиховог имена пише се страним ћириличним или страним латиничним писмом, „паралелну одредницу“ чини етимолошки ћирилични облик и облик транслитерован на латиницу, уколико је етимолошки облик имена ћирилични, али не српски (пример 2) или фонетски ћирилични српски и етимолошки латинични облик (пример 3).

Ова фаза пројекта трајала је дуже него што је предвиђено; закључена је 2010. године и резултирала је корпусом од преко 54.000 библиографских записа који је чинио основ првобитне нормативне датотеке личних имена CONOR.SR инсталиране 2013. године.

Да би се применила позитивна искуства словеначких колега, као модел увођења нормативне контроле у систем узајамне каталогизације преузет је словеначки пројекат уз извесне модификације. Први корак било је инсталирање програмске опреме COBISS3/Katalogizacija. Услов за формирање иницијалне нормативне датотеке био је да све библиотеке у систему пређу на ову верзију програма. Организована је и едукација редактора за ажурирање нормативних записа и утврђивање облика и структуре паралелних одредница. У окружењу COBISS3 током 2013. године испуњене су све претпоставке за почетак редакције иницијално аутоматски креираних нормативних записа.

На основу искустава словеначких колега одлучено је да се изврши редакција свих нормативних записа пре повезивања нормативне и библиографске базе података и уследила је обука главних редактора. Словеначки пример аутоматског повезивања нормативне и библиографске базе података без претходне редакције нормативних записа није се показао као ефикасан нити економичан. Нередиговани нормативни записи нису допринели квалитету библиографске базе нити олакшали процес каталогизације. Та ситуација је наметнула потребу истовременог редиговања нормативног записа и свих библиографских записа повезаних са њим. Што се тиче организационог дела рада на нормативној датотеци личних имена, формиран је редакторски тим од шест редактора (по два редактора из Народне библиотеке Србије, Библиотеке Матице српске и Универзитетске библиотеке „Светозар

Марковић“) који је и највиши ниво у даљој реализацији пројекта. Задаци главног редакторског тима подразумевали су прављење плана рада, утврђивање принципа редакције, подела послова и израда упутства за редакцију нормативне базе CONOR.SR.

Редакција иницијалне нормативне датотеке личних имена започета је 2015. године и ангажовани су били само главни редактори (шест редактора). Изузетно сложен посао редакције нормативних записа није се одвијао задовољавајућом брзином и постепено се редакторски тим ширио, најпре у Библиотеци Матице српске, затим Универзитетској библиотеци „Светозар Марковић“ и, на крају, у Народној библиотеци Србије. На ажурирању нормативне датотеке личних имена тренутно су ангажована 23 редактора и редигована су 8.533 нормативна записа.[7] Планирано је да се редакторски тимови шире и да се с временом повећа ефикасност рада. Израђено је и Упутство за редиговање припремне базе CONOR.SR.

Циљ пројекта израде нормативних датотека за лично име, породично име, назив колективног тела, географска имена, општих предметних одредница и јединствених наслова, како је на почетку постављен, првенствено је обезбеђивање контроле квалитета и уједначавања података што би олакшало рад каталогизаторима и класификаторима, а корисницима омогућило ефикасније претраживање и коришћење извора. То се постиже доследном употребом утврђених преферентних облика означавања термина, појмова и ентитета. Успостављањем нормативне контроле повећава се и присуство библиотека на глобалној мрежи и успешније коришћење њихових ресурса. Поред тога, циљ је био и поспешити међународну размену података и побољшати сарадњу са другим библиотекама и сродним установама. Без обзира на све позитивне аспекте пројекта, треба узети у обзир и чињеницу да у Србији и даље постоје библиотеке које ће само посредно имати користи од резултата израде нормативне датотеке будући да своје електронске каталоге одржавају и развијају у оквирима система који не подржавају размену библиографских података са библиотекама које раде под COBISS платформом. Нормативне датотеке су замишљене и као извор лексикографских и биобиблиографских података, те у том смислу могу да имају широку употребу. Директна, непосредна размена података јесте најшири оквир нормативне контроле, али не треба негирати и значај коришћења нормативних датотека као извора података.

[7] Подаци на дан 1. 8. 2016.

С обзиром на то да нормативна база личних имена CONOR.RS почива на међународно признатим стандардима[8] и да је тенденција да се пројекат шири и на остале приступне тачке библиографских записа, дугорочан циљ је, свакако, партиципирање у виртуелној међународној нормативној датотеци.

4. Нормативна контрола у 21. веку

Растући тренд отварања података и омогућавање приступа подацима и њиховом коришћењу и размени широкој јавности односи се на многе националне и међународне системе, и библиотеке су позване да се укључе у ову агенду. Први корак је превазилажење граница библиотекарске заједнице и прихватања комуникације методом отворених повезаних података. Шире посматрано, библиотеке усвајају нове начине остваривања једног од својих основних циљева – ширења знања. У новонасталом активном окружењу то је било могуће постићи само новом структуром података, прилагођеном савременим начинима комуникације које оне већ поседују и нуде на коришћење. То није био мали корак. Постојећи стандарди који су се тицали каталогизације, како наводи Д. Хилман и др., били су намењени корисницима и односили су се на текстуалне документе или њихову представу на интернету у облику документа (Hillmann et al., 2010). Промене су се односиле на прилагођавање библиотека на интеракцију у релацији рачунар-рачунар, али на другачији начин од оног који смо познавали, односно интеракцију између библиотечке заједнице и оне која то није, а размена података међу њима била је све неопходнија. Семантички веб и концепт отворених

[8] Међу најзначајнијим стратешким документима су: *Упутства за нормативне и упутне јединице (GARE – Guidelines for Authority and Reference Entries)*, објавила Ифла 1984. године; *Упутства за предметне одреднице и упутне јединице (GSARE – Guidelines for Subject Authority and Reference Entries)*, објавила Ифла 1993. године и формат *UNIMARC/Нормативни подаци (UNIMARC/Authorities)* такође Ифлино издање из 1991. године. Касније су објављена још три документа на којима се заснива нормативна контрола: *Функционални захтеви за библиографске записе (FRBR – Functional Requirements for Bibliographic Records)* 1998. године; *Функционални захтеви за нормативне податке (FRAD – Functional Requirements for Authority Data)* 2009. године и *Функционални захтеви и нумерисање нормативних записа (FRANAR – Functional Requirement and Numbering of Authority Records)* 2009. године.

повезаних података постао је нова стварност и платформа којој су библиотеке, прилагођавајући се, имале шта да понуде. Контролисане листе термина, класификационе шеме, тезауруси, листе предметних одредница и нормативне датотеке, као библиотечка артефакта, били су погодни за структурирање и адекватни за размену података и ван библиотечке заједнице.

Усвајањем RDF модела (*Resource Description Framework*), основног оквира за дефинисање, употребу и размену метаподатака који је повећао интероперабилност података којима библиотеке располажу, оне преузимају активнију улогу нудећи своје ресурсе веб корисницима и узимају учешће у семантичком вебу. Тешко је било променити свест да је то једини ваљани избор и да би у супротном библиотеке биле игнорисане у свету информација. Ка тој промени библиотеке су кренуле, како Г. Бирн и Л. Годард наводе, развијањем RDA (*Resource Description and Access*), FRBR (*Functional Requirements for Bibliographic Records*) и FRAD (F*unctional Requirements for Authority Data*) (Byrne and Goddard, 2010). Неизбежно је сагласити се са истим ауторкама да је RDA покренуо бројна питања и изазвао неповерење библиотекара према новом концепту, нарочито међу каталогизаторима. Сврха развијања поменутих модела било је отварање колекција на вебу и истрајавање библиотека у одржању својих фундаменталних принципа у измењеном окружењу.

Нормативне датотеке које библиотеке креирају биле су први озбиљни адути за учешће библиотека у интеграцији информација и лакше проналажење бољих путева лоцирања и коришћења информација под окриљем семантичког веба. Оне су биле пут ка откривању библиотечких колекција и у машински читљивом облику, уз активну улогу и двосмерни проток информација. Онлајн јавни каталози су одавно задовољавали овакву потребу, али сада треба да одговоре на захтев да понуђене информације буду у машински читљивом облику, присутне на интернету у стандардизованом облику који се може обогаћивати и лако повезивати. У том контексту, од 2006. године ISNI као међународни стандард једнозначно и недвосмислено идентификује имена стваралаца и називе колективних тела, а њих је тренутно 9,12 милиона[9] и повезан је на вебу са VIAF-ом. VIAF је један од првих покушаја у домену библиотечких ресурса да се употребом напредне технологије повежу нормативни записи више библиотека у свету што

[9] Подаци на дан 7. 8. 2016.

је као крајњи циљ имало стварање јавно доступног интернет сервиса и унапређивање нормативне контроле на националним нивоима, то јест, унапређивање рада библиотека, али и њихово повезивање са широм заједницом. Потом, Британска библиотека је своју Британску националну библиографију 2011. године почела да развија под моделом отворених повезаних података тако што је у одређено поље нормативног записа укључила локални URI, а слично је учинила и Конгресна библиотека, као и Француска национална библиотека која нормативним записима додељује трајни идентификатор изведен из идентификационог броја записа. Исто се односи и на друге европске националне библиотеке попут немачке, шпанске, мађарске итд. То су примери све већег присуства библиотека на интернету и њихове употребне вредности и својим традиционалним корисницима и веб заједници. Тако нормативни записи личних имена, као сегменти база података које имају карактеристике биобиблиографских лексикона, широкој заједници корисника нуде податке о личном имену особе која је повезана са неким интелектуалним или уметничким садржајем, алтернативним облицима имена укључујући и облике имена на другим језицима, податке о години рођења (и смрти), остале биографске податке и својеврстан извадак из библиографије, што илуструје пример 4.

Пример 4

```
001  <a>c - исправљени запис <b>x - нормативни запис
<c>a - лично име

100  <b>a - усвојена <c>srp - српски <g>cb - ћирилица - српска
101  <a>srp - српски

102  <a>srb - Србија <b>cs - Централна Србија <a>bih -
Босна и Херцеговина <b> rs - Република Српска

106  <a>0 - може да се користи и као предметна приступна тачка
120  <b>a - приступна тачка за једну особу <a>a - женски
152  <a>PPIAK - Правилник и приручник за израдбу
абецедних каталога 19011 <a>1951 <b>01 <c>31

200  1 <7>cb - ћирилица - српска <a>Радуловић <b>Здравка
<f>1951- <r>04130
```

```
200  1 <7>ba - латиница <a>Radulović <b>Zdravka <f>1951- <r>04130

340 <a> Библиограф саветник. На Одсјеку за Општу књижевност
и библиотекарство при Филозофском факултету у Сарајеву
стекла диплому Дипломираног библиотекара и дипломираног
компаративисте. На истом Одсјеку завршила постдипломске студије
и магистрирала 1988. г. Радила на Катедри за библиотекарство
(Одсјек за Општу књижевност и библиотекарство) у Сарајеву,
у Библиографском одјељењу Централне народне библиотеке СР
Црне Горе "Ђурђе Црнојевић", од 1997. г. радила
у Библиографском одељењу Народне библиотеке Србије чији је
начелник била од 2005.-2015. г.
Добитница награде за изузетна остварења
у области библиографије "Душан Панковић‘‘.

810 <a> Библиографија часописа "Библиотекар":
(1948-1997) / Здравка Радуловић, Долорес Калођера-Петровић

810  <a>Evropske integracije [Elektronski izvor] :
bibliografija monografskih publikacija i
članaka u serijskim publikacijama : 1995-2008 /
[izrada bibliografije Zdravka Radulović]. - 2009

810  <a>Библиографија часописа "Дабро-босански Источник" :
(1887-1911) / Здравка Радуловић. - 2010
```

Нормативне датотеке које библиотеке развијају у оквиру својих система, најшире гледано, имају двојак значај. Унапређују делатност самих библиотека унутар система и шире њихов утицај на целокупну информациону заједницу. Корист коју библиотеке имају од нормативних датотека унутар библиотечке заједнице односи се на унапређење пословања, побољшање услуга које нуде, проширивање циљне групе којој се обраћају и тиме преузимају активну улогу у окружењу. Из такве улоге произилази и други аспект значаја који превазилази границе једног сервиса и укључује се у општу размену информација и савремени комуникациони модел.

5. Закључак

Концепт отворених података примењен на библиотеке као јавне установе подразумева да подаци које оне креирају и похрањују буду јавно доступни, да њихова употреба буде слободна и да могу поново да се користе. Када говоримо о библиотечким каталозима и метаподацима садржаним у њима, остваривање тог концепта није било могуће одржавањем библиографских база података које су релационе базе и било је неопходно повећати интероперабилност података и учинити их доступним преко веба. То је постигнуто укључивањем нормативних датотека у постојеће електронске каталоге. Тако структурирани подаци постали су део концепта отворених података чиме је оснажено присуство библиотека на глобалној мрежи, проширен је круг корисника библиотека и пружен је одговор савременом добу које „доба глобалног библиотекарства, заснованог на сарадњи и размени ресурса и информација и партнерству на свим нивоима“ (Билбија, 2015). Сакупљени делови информација се комбинују, ширењем информација, на нов начин изграђује се партнерски однос са корисницима и пружа квалитетнија услуга. Основни принцип универзалне библиографске контроле је заједничко коришћење информација. Савремене технологије пружају нове начине остваривања тог принципа, а нормативним датотекама библиотеке обезбеђују релевантност података на глобалном нивоу, ван националних оквира и ван библиотечке заједнице. Такође, омогућавају размену података међу технологијама. Ка том циљу су усмерени и напори српске библиотекарске заједнице. Општи циљ је акумулирање знања, његово систематизовање и потпуно отворено, заједничко коришћење целокупног људског знања.

Литература

Angjeli, Anila and Andrew Mac Ewan and Vincent Boulet. “ISNI and VIAF – Transforming ways of trustfully consolidating identities”. Рад представљен на: *IFLA WLIC 2014 – Lyon – Libraries, Citizens, Societies: Confluence for Knowledge in Session 86 – Cataloguing with Bibliography, Classification & Indexing and UNIMARC strategic Programme* (2014). In: *IFLA WLIC 2014, 16-22 August 2014, Lyon, France.* Преузето 22.7.2016, library.ifla.org/985/1/086-angjeli-en.pdf.

Byrne, Gillian and Lisa Goddard. “The strongest link: Libraries and linked data”. *D-Lib Magazine* Vol. 16 no. 11/12 (2010). Преузето 7.8.2016, http://www.dlib.org/dlib/november10/byrne/11byrne.html.

Hillmann, Diane, Karen Coyle, Jon Phipps and Gordon Dunsire. “RDA vocabularies: process, outcome, use”. *D-Lib magazine* Vol. 16 no. 1/2 (2010).

Seljak, Martaetal. “Vzpostavitev normativne kontrole v knjižničnem informacijskem sistemu COBISS.SI, Slovenija”. *Organizacija znanja* Vol. 9 no. 2: 1–16 (2004). doi:10.3359/oz0402037.

UBC, IFLA International Office for and UNESCO. “Guidelines for the National Bibliographic Agency and the National Bibliography”. United Nations Educational, Scientific and Cultural Organization, 1979. Преузето 22.7.2016, nesdoc.unesco.org/images/0004/000486/048658eo.pdf.

Билбија, Биљана. *Основе библиотекарства 2.* измијењено и допуњено изд. Бањалука: Народна и универзитетска библиотека Републике Српске, 2015.

Леива-Медерос, Амед и други. “AUTHORIS – инструмент за нормативну контролу на семантичком вебу”. *Гласник Народне библиотеке Србије 2014/2015* (2015): 35–53

Изложба „Дигитална ћирилица“

Милена Костић
mkostic@unilib.rs

Весна Вуксан
vuksan@unilib.rs

Зоран Бајин
bajin@unilib.bg.ac.rs

Универзитетска библиотека „Светозар Марковић“

РАД ПРИМЉЕН: 15. јул 2016.
РАД ПРИХВАЋЕН: 10. децембар 2016.

Обележавајући своју деведесету годишњицу, која је истовремено и Дан библиотеке и Дан Ћирила и Методија и словенске писмености и културе, Универзитетска библиотека „Светозар Марковић“ приредила је физичку, виртуелну и веб-изложбу „Дигитална ћирилица“која је трајала од 24. маја до 1. јула 2016. Аутори изложбе су библиотекари Универзитетске библиотеке Весна Вуксан, Милена Костић и Зоран Бајин. На свечаном отварању изложбе 24. маја 2016. године у Карнегијевом холу Библиотеке присутнима су се обратили проф. др Нада Ковачевић, проректор за наставу Универзитета у Београду, проф. др Александар Јерков, управник Универзитетске библиотеке „Светозар Марковић”, Тамара Бутиган Вучај, заменица управника Народне библиотеке Србије, др Адам Софронијевић, заменик управника Универзитетске библиотеке „Светозар Марковић” и путем Скајпа проф. др Гинтер Милбергер са Универзитета у Инсбруку и Клаус Гравенхорст из компаније „CCS Content Conversion Specialists GmbH“[1] у Хамбургу која се пре свега бави дигиталним библиотекама и архивима. Снимак са свечаног отварања изложбе можете погледати на линку: http://media.rcub.bg.ac.rs/?p=5259.

1. Изложба у Библиотеци

Изложба у самој Библиотеци, поред одабраних књига о раду и значају Ћирила и Методија, обухватила је и преводе дела словенских

[1] http://content-conversion.com/

књижевности на српски језик, објављиване од средине XIX века до почетка Првог светског рата. Ови преводи допринели су оживљавању културних веза између Срба и осталих словенских народа и језика, чије су заједништво пресудно учврстили Ћирило и Методије хиљаду година раније. У оквиру изложбе представљен је дигитални лисни каталог доступан на сајту Универзитетске библиотеке: http://www.unilib.rs/pretraga/katalog/.

2. Виртуелна изложба

Идући у корак са најновијим технологијама у библиотекарству, и водећи рачуна о заштити ретких књига, Универзитетска библиотека је под покровитељством Министарства културе и информисања набавила Magic Box, уређај који омогућава виртуелно, али по осећају готово стварно прелиставање драгоцене и заштићене библиотечке грађе. Magic Box, који се налази у приземљу библиотеке, премијерно је представљен јавности на свечаном отварању изложбе. У магичној витрини на Дан библиотеке, физички је било изложено Четворојеванђеље, као изузетан примерак рукописних књига створених захваљујући писму које су словенским народима оставили Ћирило и Методије. Виртуелно (дигитализовано) Четворојеванђеље може се „прелистати“ на интерактивном екрану, као и репрезентативни примерци грађе коју Библиотека поседује: стари ћирилски рукописи, историјске новине, фотографије, избор из књига изложених у витринама, видео клипови и 3Д објекти.

На овом линку http://www.unilib.rs/pretraga/katalog/ можете погледати презентацију Клауса Гравенхорста из компаније „CCS Content Conversion Specialists GmbH“ из Хамбурга, која производи Magic Box. У презентацији је представљен програм doc Works који унапређује рад са дигиталним документима, пре свега могућношћу да се креирају METS/ALTO фајлови. Овај програм ће Библиотека добити у 2016. години кроз пројекат „Нови хоризонт дигитализације у Србији” под покровитељством Министарства културе и информисања. У презентацији је такође представљен Magic Box, дигитална изложбена витрина, која приказује METS/ALTO фајлове.

3. Веб-изложба

Веб-изложба нуди приступ дигитализованим ћирилским рукописима и књигама о раду и значају Ћирила и Методија, као и преводима дела словенских књижевности на српски језик који су објављивани од средине XIX века до почетка Првог светског рата, а доступна је на сајту Универзитетске библиотеке на линку: http://www.unilib.rs/sadrzaji/virtuelne-izlozbe/. У наставку дајемо опис дигитализоване грађе.

3.1 Ћирилски рукописи

Универзитетска библиотека „Светозар Марковић“ виртуелно је изложила седам репрезентативних рукописних књига из своје збирке ћирилских рукописа. Јединствене уметничке опреме и лепих писарских потеза, рукопис Четворојеванђеља представљен је и непосредно у изложбеном простору дигиталног уређаја. Место и време настанка изложеног Четворојеванђеља, које се одликује калиграфском вештином писара и илустрацијама са представама четворице јеванђелиста, нису још тачно утврђени. На корицама се налази запис којим се сведочи да се књига 1685. године налазила у поседу јеромонаха Пахомија „од Црне Реке“ — манастира Црне Реке. Поред Четворојеванђеља, виртуелно су представљени: Панагирик писара Венијамина из манастира Свете Тројице код Пљеваља из 1595. године; Минеј за јануар са Службом Светом Сави од Теодосија, настао крајем XVI или почетком XVII века; Псалтир с последовањем, писан 1652. године, са додацима из XVIII века, на коме се налази запис „Доситеја ђакона хоповскога“ – вероватно Доситеја Обрадовића; Служабник из треће четвртине XVI века, у коме се налази и Кратка служба Светом Симеону; Барањски препис Душановог законика, настао у првој половини XVI века, коме су, као и другим преписима Душановог законика, прикључени скраћена Синтагма Матије Властара и Јустинијанов зборник; као и зборник древних повести са саставом индијских поучних прича – Поведаније о древних и повјестех са Стефанитом и Ихнилатом – Ћоровић 31.

3.2 Књиге о Ћирилу и Методију

У Библиотеци су физички биле изложене одабране књиге о животу и деловању Ћирила и Методија. Из њих су за дигитализацију издвојена

три наслова: српски превод дела Теофилакта Орхидског о Ћирилу и Методију (Истинаја повест о Кириле и Методији и о изобретенији богоданих кирилических и славесних називајемих писмен), објављен у Будиму 1823. године; Хиљадугодишња светковина Св. Ћирила и Методија, са говором Димитрија Матића и кратком Методијевом биографијом од Јанка Шафарика, објављена у Београду 1863. године; као и илустрована Geschichte der Heiligen Slaven-Apostel Cyrill und Method (Историја светих словенских апостола Ћирила и Методија) чешког свештеника и историчара Јана Билија, објављена у Прагу те исте, јубиларне године.

3.3 Преводи дела словенских књижевности

Највећи део физичке изложбе чинили су преводи дела словенских књижевности на српски језик, објављивани од средине XIX века до почетка Првог светског рата. Из њих је за дигитализацију издвојено седам наслова. Од превода дела руске књижевности одабрани су најстарији: Капетанова ћерка Александра Пушкина, објављена 1849, Шињел Николаја Гогоља, објављен у Мостару 1902. године, и Хаџи Мурат, последње белетристичко дело Лава Толстоја, чији се превод на српски појавио уочи самог избијања светског рата. Од превода дела пољске књижевности одабрани су Козачка освета Михаила Чајковског, објављена у Београду 1854, и Приповетке Хенрика Сјенкјевича, објављене у Новом Саду 1887. године. Од превода дела чешке књижевности одабран је Сеоски роман Каролине Свјетле, објављен у Београду 1879, а од превода бугарске књижевности сатирични и забавни Баја-Гање Алека Константинова из 1907. године.

Догађај је био медијски испраћен на телевизији, радију и интернет порталима и укупно је било 26 медијских појављивања. Већина портала пренела је вест[2] Радио Телевизије Србије. Садржај вести варирао је од оних које су фокус усмериле на опис изложбе „Дигитална ћирилица“ до оних које су се више бавиле описом Magic Box-а.

[2] http://www.rts.rs/page/stories/ci/story/124/drustvo/2328542/vremeplov-kroz-istoriju-uz-digitalnu-vitrinu.html

Приказ летње школе о претраживању великих повезаних скупова података заснованом на кључним речима „Keyword search in Big Linked Data“

Марија Радојичић
Ранка Станковић
Универзитет у Београду
Рударско–геолошки факултет

Себастијан Каплар
Универзитет у Новом Саду
Факултет техничких наука

РАД ПРИМЉЕН: 9. април 2017.
РАД ПРИХВАЋЕН: 8. мај 2017.

У оквиру COST акције[1] IC1302 под називом KEYSTONE (semantic KEYword-based Search on sTructured data sOurcEs)[2] одржана је друга истраживачка летња школа под називом „Keyword search in Big Linked Data“ у периоду од 18. до 22. јула 2016. године у граду Сантјаго де Компостела у Шпанији. Циљ школе био је да се учесницима представе актуелне теме у области која се брзо развија, а односи се на претрагу великих количина повезаних података (Big Linked Data) засновану на кључним речима. Школа је намењена углавном дипломираним студентима и постдипломцима на почетку њихове академске каријере, а организатор је био Центар за истраживања информационих технологија[3] Универзитета Сантјаго де Компостела.[4]

Основна идеја летње школе била је анализа и управљање великим повезаним подацима што је појединачно укључило теме: велики подаци (Big Data), повезани подаци (Linked Data), обрада природног језика (NLP, Natural Language Processing), семантички веб (Semantic Web) и проналажење информација (IR, Information Retrieval). Осам истакнутих

[1] European Coopeartion in Science and Technology. www.cost.eu/COST_Actions
[2] http://www.cost.eu/COST_Actions/ict/IC1302
[3] Centro de Investigación en Tecnoloxías da Información (CiTIUS), https://citius.usc.es/
[4] Universidade de Santiago de Compostela, http://www.usc.es/

предавача из ових области припремили су интересантан материјал, који је јавно доступан[5]. Летњу школу похађало је 38 учесника из тринаест земаља са три континента – Европе, Америке и Африке. Летња школа је организована тако да су учесници током преподнева имали прилику да чују истакнуте предаваче са различитих европских универзитета, док су у оквиру поподневних радионица на практичним примерима могли да испробају научено. Такође, учесници су имали прилику да чују и искуства представника Hewlett Packard-а као и фирме E˜Xenio о теми „Keyword search in Big Linked Data“.

Лаура По, предавач на одсеку за инжењерство Универзитета Модена и Рађо Емилија[6] из Италије, говорила је о истраживању, визуелизацији и постављању упита над великим скуповима повезаних података. Пре свега учесницима је објашњено шта су то повезани подаци, а шта отворени подаци, како се они публикују, каква је њихова улога у семантичком вебу, након чега се говорило о истраживању и постављању упита над великим скуповима повезаних података. Посебна пажња посвећена је SPARQL[7] језику који се користи за постављање упита у циљу претраживања отворених повезаних података. Представљен је концепт чворишта за податке (DataHub), претраживање метаподатака о расположивим скуповима података, начини и формати преузимања скупова података, као и коришћење SPARQL приступних тачака (SPARQL endpoint) уколико су наведени уз конкретан скуп. Истакнута је важност визуелизације повезаних отворених података и демонстриран алат LODeX са конкретним примерима упита и резултујућих интерактивних графова.

У наставку летње школе ученици су се упознали са основним појмовима и техникама везаним за проналажење информација (IR). Кроз свеобухватно и интересантно предавање Михаила Лупуа са Техничког универзитета у Бечу[8] учесницима су предочени различити начини евалуације, као и кључни критеријуми за оцену IR. Посебна пажња посвећена је примени статистичких метода за проверу добијених резултата. О тој теми говорио је и Сергеј Зер са Универзитета Саутхемптон[9] који је својим предавањем „Collective Intelligence: Crowd-

[5] https://eventos.citius.usc.es/keystone.school/slides.html

[6] Università degli Studi di Modena e Reggio Emilia (University of Modena and Reggio Emilia), http://www.unimore.it/

[7] SPARQL Protocol and RDF Query Language

[8] Technische Universität Wien (TUW), www.tuwien.ac.at

[9] University of Southampton, www.southampton.ac.uk

sourcing Groundtruth Data for Large Scale Evaluation in Information Retrieval“, држао пажњу присутних од првог до последњег минута, покренувши их притом да се укључе у дискусију тако што је приказао занимљиве примере.

Садржајно предавање Женевјев Варга-Солар, истраживача из Француског националног центра за научна истраживања,[10] значајно је употпунило слику о досадашњим достигнућима по питању чувања и обраде великих повезаних података. Она је посебну пажњу посветила трендовима у анализи великих података, истраживању података (data mining) и науци о подацима (data science), као и о специфичностима дистрибуираног моделирања, складиштења, предиктивне аналитике, кластеровања, обраде и истраживања токова података (stream processing & mining), декларативним језицима.

Елена Демидова са Универзитета Саутхемптон упознала је полазнике летње школе са интерактивном претрагом великих структурираних скупова података заснованом на кључним речима. Претрага структурираних података путем кључних речи, наспрам класичних упитних језика (SQL), захтева претходну припрему података (индексирање) и превођење информационих потреба корисника у погодан облик. Добра страна оваквог приступа је могућност постављања упита чак и када шема базе података није позната, али је недостатак непрецизност интерпретације информационе потребе корисника јер се полазни упит изражен кључним речима преводи у (највероватнији) структурирани упит. Конструкција упита за велике базе података, на пример Freebase са преко 22 милиона ентитета, 350 милиона чињеница, 7.500 релационих табела и око 100 домена, захтева ефикасну и прошириву интерактивну конструкцију упита која је представљена теоретски и кроз примере.

Током летње школе Ранка Станковић, професорка Рударско–геолошког факултета Универзитета у Београду, одржала је предавање о језичким ресурсима. Посебну пажњу професорка Станковић посветила је проширивању упита и семантичком означавању као начинима за унапређење претраживања у смислу повећања одзива без губитка прецизности. Током предавања полазници су упознати са различитим алатима за језичку обраду. У оквиру радионице професорка Станковић

[10] Centre national de la recherche scientifique, French Council of Scientific Research (CNRS), http://www.cnrs.fr

представила је до сада развијене ресурсе за српски језик и демонстрирала њихову примену.

Мауро Драгони из Фондације Бруно Кеслер[11] последњег дана предвиђеног за предавања истакао је значај сагледавања неког документа из различитих углова и приказао неколико студија случаја које су допринеле дубљем разумевању представљених концепата.

На самом крају летње школе уследио је можда најзанимљивији део када су се учесници здружили у групе од три до четири члана како би учествовали у такмичењу Hackaton. Задатак који су учесници добили захтевао је како програмерске вештине тако и примену алата и знања представљених за време летње школе. Током такмичења коришћена су два опште позната скупа повезаних података за решавање постављених проблема: DBPedia и GeoNames. DBPedia је централни репозиторијум повезаних података, екстрахован из Википедије, који описује више од 4,5 милиона ентитета класификованих у конзистентну онтологију од којих су око 1.445.000 ентитета за особе, 735.000 за места, 411.000 за уметничка дела, 241.000 за организације. Постоје верзије за 125 различитих језика, при чему је онтологија за енглески језик највећа. Осим линкова ка сликама, спољним странама, ентитетима су придружене категорије из Википедије и YAGO онтологије, што омогућава постављање SPARQL упита над подацима изведеним из Википедије. Бројни скупови података и онтологије из репозиторијума DBPedia могу се преузети или им се може директно поступати онлајн коришћењем SPARQL приступне тачке`http://dbpedia.org/sparql`, претраживати по кључним речима коришћењем алата DBPedia Lookup[12] или се могу преузети и користити као локални повезани подаци.

Географска база података GeoNames[13] садржи преко десет милиона географских назива и девет милиона јединствених географских ентитета, класификованих у девет класа, означених са 645 маркера. Ове класе и маркери су описани GeoNames онтологијом[14] и предефинисаним кодовима.[15] Сваки елемент у GeoNames има URI и одговарајући документ са RDF XML подацима. Тако је, на пример, елемент за Београд има URI `http://www.geonames.org/792680/` и одговарајући документ `http://www.geonames.org/792680/about.rdf` у ком је, осим назива на

[11] Fondazione Bruno Kessler, `www.fbk.eu`
[12] `http://wiki.dbpedia.org/projects/dbpedia-lookup`
[13] `http://www.geonames.org/`
[14] `http://www.geonames.org/ontology/documentation.html`
[15] `http://www.geonames.org/export/codes.html`

енглеском језику, Београд описан на још 67 језика. GeoNames елементи су међусобно повезани коришћењем три типа географских релација: подређени, у смислу административних потцелина, потом суседни, на пример суседне државе и просторно блиски, на пример насеља које се налазе на малој удаљености. Тако се свим регионима у Црној Гори може приступити када се на основни њен http://www.geonames.org/3194884 дода суфикс contains.rdf.

Податке из GeoNames базе могуће је преузети и користити локално, а онлајн приступ је могућ путем кореног чвора његове хијерархије http://sws.geonames.org/6295630/about.rdf и праћењем *contains* релација, или коришћењем веб-сервиса заснованог на кључним речима.[16] Коначно, може се видети да постоје линкови од GeoNames елемената ка Википедији и DBPedia-ји.

Полазници су добили задатак да користећи репозиторијум DBPedia и GeoNames направе веб-апликацију за: 1) проналажење административних области погођених задатим ураганом, 2) проналажење свих урагана који су погодили задате административне области и 3) процењивање како су одређене области биле припремљене за задати ураган.

1. У првом сценарију корисник треба да зада кључне речи којима идентификује ураган, што укључује његово име (Катрина, Емили итд.) и опционо годину (2005, 2006. итд.). Текст о урагану треба пронаћи у репозиторијуму DBPedia, потом га издвојити, при чему претрага треба да укључи апстракт и текст који описује погођене области. Из исечака текста потом треба издвојити имена административних области коришћењем GeoNames базе и Standford NER[17] алата за препознавање именованих ентитета.
2. У другом сценарију корисник кључним речима задаје административну област, а потом се GeoNames база користи за проналажење насеља у задатој области, након чега се претражује DBPedia како би се идентификовали урагани који су погодили та насеља, односно област.
3. Трећи део је подразумевао одређивање нумеричког индикатора којим се оцењује припремљеност неке административне области за ураган коришћењем података из DBPedia-је (жртве, штета, брзина ветра, трајање, тип и слично) и броја становника у угроженим

[16] http://www.geonames.org/export/geonames-search.html
[17] http://nlp.stanford.edu/software/CRF-NER.shtml

областима пронађеним у претходном упиту. Потребно је генерисати рангирање засновано на предложеном индикатору.

Једна од стратегија српског тима за решавање датог проблема била је креирање апликације која кроз програмски језик Python користи све расположиве отворене ресурсе како би приступила датим подацима и обрадила их. Идеја је била да након што корисник одабере ураган о коме жели да сазна више информација, систем шаље SPARQL упит ка отвореном API-ју DBPedia-је који је издвајао текст апстракта и текст који описује погођене области. Након тога из резултата претходног упита су извучене погођене области и слао се нови упит ка GeoNames бази података који је враћао информације о свим погођеним насељима за дату област. Из добијених ресурса издвојене су информације попут броја становника, броја жртава, материјалне штете, брзине ветра и слично, на основу чега су се одредили нумерички индикатори који су приказали спремност неке одређене административне области за ураган. Ти нумерички индикатори су предвиђени за креирање математичког модела и његов визуелни приказ и интерпретацију.

Последњег дана летње школе, који је протекао у правом такмичарском духу и преданом рада учесника, представљена су решења екипа. Најуспешније решење је направила екипа студената мастер и докторских студија из Сарагосе, тако да су као најбоља екипа освојили вредну награду у виду стипендије за боравак и истраживање на Универзитету Сантјаго де Компостела. Евалуација система је подразумевала процену прецизности и одзива система, брзине извршавања, као и ефектности решења. Након свечаног затварања учесници су имали прилику да уживају у живописном граду Сантјаго де Компостела, као и у чарима Галиције.

Сви материјали са летње школе доступни су на адреси https://eventos.citius.usc.es/keystone.school/index.html.

УПУТСТВО ЗА АУТОРЕ

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

КАТЕГОРИЗАЦИЈА РАДОВА

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

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

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

ЕЛЕМЕНТИ РУКОПИСА

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

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

ПРАВИЛА УРЕЂИВАЊА ПРИХВАЋЕНИХ РАДОВА

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

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

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

БИБЛИОГРАФИЈА И ЦИТИРАЊЕ

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