

Impressum

FOR THE EDITOR:

Prof. Aleksandar Jerkov, PhD
University Library "Svetozar Marković"
Faculty of Philology, University of Belgrade

office@unilib.bg.ac.rs

EDITOR:

Faculty of Philology, University of Belgrade
University Library "Svetozar Marković"
Serbian Academic Library Association

EDITOR-IN-CHIEF

Prof. Cvetana Krstev, PhD
Faculty of Philology, Department for Library and Information Science

cvetana@matf.bg.ac.rs

MANAGING EDITOR:

Aleksandra Trtovac, PhD
University Library "Svetozar Marković"

aleksandra@unilib.bg.ac.rs

EDITOR OF ONLINE EDITION:

Jelena Andonovski
University Library "Svetozar Marković"

andonovski@unilib.bg.ac.rs

EDITORIAL BOARD :

Prof. Aleksandra Vraneš, PhD, Prof. Aleksandar Jerkov, PhD, Prof. Biljana Dojčinović, PhD, *Faculty of Philology, University of Belgrade*; Prof. Elisabeth Burr, PhD, *Institut für Romanistik, Universität Leipzig*; Prof. Vladan Devedžić, PhD, *Faculty of Organization Sciences, University of Belgrade*; prof. Milena Dobrova, PhD, *Faculty of Media and Knowledge Sciences, University of Malta*; Tomaž Erjavec, PhD, *Department of Knowledge Technologies, Jozef Stefan Institute, Ljubljana*; Prof. Svetla Koeva, PhD, *Institute for Bulgarian Language, Bulgarian Academy of Sciences*; Prof. Denis Maurel, PhD, prof. Agata Savary, PhD, *Université Francois Rabelais de Tours*; Prof. Ivan Obradović, PhD, *Faculty of Mining and Geology, University of Belgrade*; Prof. Gordana Pavlović Lažetić, PhD, prof. Duško Vitas, PhD, *Faculty of Mathematics, University of Belgrade*; Prof. Katerina Zdravkova, PhD, *Faculty of Computer Science and Engineering, Ss. Cyril and Methodius University, Skopje*

ISSN 1450-9687 (print edition)
ISSN 2217-9461 (online edition)

Belgrade, Vol. 17, No. 2, December 2017

WEB PORTAL:

Vukosav Sredojević

University Library "Svetozar Marković"

LECTOR FOR ENGLISH:

Zoran Ristović, PhD

Elementary School "Mito Igumanović", Kosjerić

LECTORS FOR SERBIAN:

Marija Pantić, Marko Vitas

DESIGN AND PREPRESS:

Aleksandar Milošević, Branislava Šandrih

and **Infotheca** Team

REDACTOR OF REFERENCES AND UDC:

Nataša Dakić

University Library "Svetozar Marković"

DOI REDACTOR:

Miloš Utvić, PhD

Faculty of Philology, University of Belgrade

JOURNAL REDACTION:

Journal Infotheca

11000 Belgrade, Bulevar kralja Aleksandra 71

+381 11 3370-211

infotheca@unilib.rs

PRINTED BY:

Dunav

Belgrade

Journal is published twice a year

Contents

Scientific papers

Cvetana Krstev and Agata Savary Games on Multiword Expressions for Community Building	7
Aleksandra Tomašević, Biljana Lazić, Dalibor Vorkapić, Mihailo Škorić, Ljiljana Kolonja The Use of the Omeka Platform for Digital Li- braries in the Field of Mining	26
Biljana Pajić eLibrary of the Goethe-Institut	50
Marija Pantić Adjectival Attributes with the Nouns “čovjek”, “žena”, “muškarac” and “muž”	66

Professional papers

Andela Stošić, Marina Milošević, Sandra Spasić, Dragica Dragosav, Ana Đorđević and Ines Stanić Working on the Multimedia Document “Al’ se nekad dobro jelo”	97
--	----

Reviews

Dragana Stolić Conference Proceedings “Legal Aspects of Digi- tisation of Cultural Heritage”	119
Jelena Andonovski, Nataša Dakić 83 rd IFLA World Library and Information Congress124	

Games on Multiword Expressions for Community Building

UDC 811.163.41'37:794

DOI 10.18485/infotheca.2017.17.2.1

ABSTRACT: In this paper we present games focusing on multiword expressions (MWEs) – word combinations that exhibit unexpected lexical, syntactic, semantic, pragmatic and/or statistical properties. These games were based on a multilingual collection of MWEs contributed by the members of a scientific network specialized in natural language processing. We describe the data collection process and their exploitation for a cross-lingual alignment game of MWEs with common meanings, a pantomime and quizzes based on idioms referring to nationalities. With these games the linguistic variety and richness of the PARSEME action members were leveraged in order to increase the networking effect.

KEYWORDS: multiword expressions, idioms, games, lexical resources.

PAPER SUBMITTED: 27 October 2017

PAPER ACCEPTED: 25 November 2017

Cvetana Krstev

cvetana@matf.bg.ac.rs

University of Belgrade

Faculty of Philology, Serbia

Agata Savary

agata.savary@univ-tours.fr

François Rabelais University

Tours

*Laboratoire d'informatique,
France*

1 Introduction and motivation

Learning through play is a renowned principle in education, as a means of effective development of social, cognitive and emotional skills, mainly in children. In scientific research the so-called serious games and games with a purpose, often based on data crowdsourcing, can be used to turn user's entertainment into a profit for a particular research task. In this paper we address the use of games on the periphery of scientific research: in community building. We report on efforts put to organize games for the COST action PARSEME,¹ which is a European scientific network in the domain

¹ PARSEME – PARSing and Multi-word Expressions, www.parseme.eu

of Natural Language Processing (NLP). It is dedicated to investigating interactions between parsing (automatic syntactic and/or semantic analysis of natural language utterances) and multiword expressions (MWEs). MWEs are word combinations which exhibit unexpected lexical, syntactic, semantic, pragmatic and/or statistical properties. Prominent examples of MWEs are idioms, such as *by and large*, *a hot dog*, or *to pull one’s leg*. Their most pervasive feature is their semantic non-compositionality, i.e. the fact that their meaning cannot be calculated from the meanings of their component words in a straightforward way. This fact makes MWEs challenging both for human non-native speakers and for NLP applications such as machine translation or information extraction. But we show that it can also be exploited for its amusing and surprising effect.

The PARSEME network, running from 2013 to 2017, gathered over 200 researchers from over 30 countries, working on as many languages. For its culminating event, which took place in September 2016 in Dubrovnik, Croatia,² multilingual examples of MWEs were collected from the participants and turned into quizzes. The underlying idea was to leverage the tremendous linguistic variety and richness of the network members, so as to achieve conviviality and increase the networking effect. As a side effect, the multilingual dataset of several hundreds of MWEs was created and is freely available.

2 State of the art

Games with a purpose (GWAPs) emerged in 2004 with the ESP image tagging game (Von Ahn and Dabbish, 2008) approximately at the same time as crowdsourcing began to be used (Howe, 2006). Actually, these two new ways of performing serious tasks have a lot in common: tasks that are still challenging for computers but are trivial (or manageable) for humans are being divided into small sub-tasks and distributed among many “workers”, that need not be specialists, to solve them. The difference between these two approaches to exploit the human brain-power lays in the way people are motivated to perform these small, usually tedious and labor-intensive, sub-tasks: in the case of crowdsourcing the motivation can be the financial gain, desire to learn something or wish to contribute to some common cause, while in the case of GWAPs the main motivation is to entertain oneself. The need to entertain human “workers/players” certainly affects the way sub-task work has to be organized and to what type of tasks games are right solutions.

² See the events page at: <https://goo.gl/uzyp3D>

Since the emergence of these two concepts voices were heard that reconsider these types of work organization. Besides the quality of the work done in this way, its ethics was also questioned. For crowdsourcing the question emerged: “Can we foresee crowd workplace in which we would want our children to participate?” (Kittur et al., 2013), while Tuite (2014) states that “Games that aim to provide only entertainment in exchange for work risk seeming exploitative, even if that was not the designer’s intention.” The same author advocates that players of a GWAP should know in advance the purpose of a game and contribute to its fulfillment of their own free will – the “entertainment” will help in doing it. Moreover, she also claims that besides doing tedious and uninteresting tasks in a human-processing unit (HPU) scenario, games should be designed for skilled professionals to perform complex and creative work, as has been already done, for instance, in the protein folding game Foldit (Cooper et al., 2010).

Crowdsourcing was successfully used in many Natural Language Processing (NLP) projects (Mitrović, 2013), especially for language resource development, and particularly for less-resourced languages for which speakers can be more easily motivated to contribute (Millour and Fort, 2017). Crowdsourcing, games with a purpose and other collective intelligence-based approaches can be used not only for building new resources for NLP, but already developed NLP resources and tools can be used to evaluate collectively built resources (Gurevych and Zesch, 2013).

In (Lafourcade et al., 2015) authors divide GWAPs into four broad categories: biological games, games with a medical purpose, GWAPs for NLP and unclassifiable games. They emphasize that GWAPs are particularly important for language resource development as they not only enable the acquisition of huge resources, but they can take into account the dynamic and evolving aspects of languages. Also, the resulting resources may incorporate those data acquired by several players on which they reached a consensus – this ascertains their soundness, something that cannot be easily obtained with traditional data acquisition. According to the inventory presented in (Lafourcade et al., 2015) most GWAPs for NLP are used for annotation tasks (anaphora relationships, coreferences, syntactic annotation, semantic disambiguation, etc.) and acquisition tasks (common sense facts, paraphrases, associations between terms and their characteristics, named entities, etc.), but also for building new resources (lexical network of terms) and validation (associations between concepts). Recent developments include no-

tably a game in MWE annotation in French.³ Finally, there are many other word or letter games that are similar to GWAPs but do not fall completely into that category (lack of consensus, no validation of players’ answers, etc.).

Games that we will present in this paper are specific in their nature. They are no full-fledged games, but were rather invented for the specific purpose of community building, and aimed to be played only once. More precisely, their main goal was to close a successful research network project in good spirits. Their side-effect achievements, besides the acquisition of resources (a multilingual inventory of MWEs), included raising the awareness of the MWE phenomenon (appropriate for a large public), and revealing stereotypes related to different nationalities which show in the idiomatic use of language.

3 Collecting the data

The input data for the games were contributed by a specific community of experts rather than via crowdsourcing. Namely, the PARSEME members were asked to fill in an online form with MWEs preferably in their mother tongues. Each MWE was to appear in the original script, transliterated to Latin script if needed, and both its literal and its idiomatic meaning were to be provided. The form consisted of 3 sections:

- highly opaque MWEs, i.e. those whose meaning is particularly hard to figure out for foreigners, as in example (1),⁴
- MWEs referring to nationalities, especially those based on positive or negative stereotypes, as in example (2),
- MWEs with general meanings, i.e. those having one of the 10 pre-selected senses: TO DIE, TO BE TIRED, TO BE STUPID, TO BE ANGRY OR IN BAD TEMPER, TO BE/FALL IN LOVE, TO DO SOMETHING USELESS, TO BE DRUNK OR FOND OF DRINKING, TO PAY ATTENTION TO UNIMPORTANT DETAILS, TO WORK HARD, TO BE LAZY; two collected examples are shown in (3)–(4).

- (1) немати длаке на језику (SR) – nemati dlake na jeziku
‘to have no hair on one’s tongue’ ⇒ TO EXPRESS ONE’S OPINION
OPENLY

³ <https://goo.gl/KhQ9yc>

⁴ Henceforth, examples will be tagged with the languages they stem from. See p. 22 for the list of language codes.

- (2) saoul comme un polonais (FR)
'as drunk as a Pole' \Rightarrow VERY DRUNK
- (3) pójść do Abrahama na piwo (PL)
'to go have a beer with Abraham' \Rightarrow TO DIE
- (4) nach dem Bettzipfel schielen (DE)
'to squint at the corner of one's bed' \Rightarrow TO BE TIRED

In this way we collected, in total, 455 MWEs from 38 contributors in 25 languages. These data were used as input for the three games presented below.⁵

4 Games on general meanings

These games were based on input data provided by users about multiword expressions for ten general senses. The idea was to choose from the users' input the most opaque examples and give them to players to guess their meaning. At the same time we wanted that all languages for which users provided the input be represented in the game. In order to give players the multiword expressions that they have never heard of, and under the assumption that close languages (languages from same families) are more likely to share the same (or similar) multiword expressions, we grouped all the collected input data into four groups: Slavic languages, Romance languages, Germanic languages and other. For instance, for the meaning TO DIE 27 MWEs were collected: 10 for Slavic languages (Bulgarian, Croatian, Macedonian, Polish, Russian and Serbian), 7 for Romance (Brazilian Portuguese, French, Portuguese, Romanian and Spanish), 4 for Germanic (Danish, English, German and Norwegian), and 10 for other languages (Greek, Hebrew, Hungarian, Lithuanian, Latvian, Maltese, Farsi and Swahili).

The players were grouped into four groups according to their mother tongue. Each group was given 12 examples chosen from the three other groups; that is, no group was presented with an example in a language that was a mother tongue of some group member. The players' task was to associate each given example with one of 4 predefined meanings (there were 3 examples for each meaning). Each example was presented with the following information: languages it comes from, the MWE in the original script, the MWE transcribed (if Latin script was not used), literal translation. For

⁵ All examples are cited in this paper as they were collected. No efforts were put into homogenising them or checking their meaning with independent experts.

instance, the Romance group obtained the following three examples for the meaning TO BE IN LOVE; TO FALL IN LOVE:

- (5) poczuć do kogoś miętę (PL)
‘to feel mint for someone’
- (6) Schmetterlinge im Bauch haben (DE)
‘to have butterflies in one’s stomach’
- (7) belécsap a villám (HU)
‘the lightning hits him’

Some of the 10 general meanings for which there were more opaque examples in various languages in input data were given as options to more than one group: TO BE ANGRY, TO BE DRUNK, TO BE USELESS, TO BE TIRED, TO DIE, TO BE STUPID.

As a complete sample of this game, we cite the dataset given to the Slavic group. It consisted of examples (8–19) to be assigned to meanings a)–d). The solution to this sample is given in Appendix 1.

Meanings:

- a) TO DIE (TO BE DEAD)
- b) TO BE TIRED
- c) TO BE STUPID
- d) TO BE ANGRY

Expressions:

- (8) estar criando malvas (ES)
‘to be growing mallows’
- (9) estar na capa da gaita (PT)
‘to be in the accordion’s case’
- (10) lóg a bele (HU),
‘his intestines are hanging’
- (11) πάω στα θυμαράκια (EL) – pao sta thimarakia
‘to go to the little thyme bushes’
- (12) qabžitlu ċ-ċinga (MT)
‘his strap got out of its place’
- (13) passer l’arme à gauche (FR)
‘to put the weapon on the left side’
- (14) har roterende fis i kasketten (DA)
‘has a rotating fart in his cap’

- (15) בַּקְלָמֶר מַחוּדָד הַכִּי הַעֲפָרוֹן לֹא (HE) – lo ha-iparon haki mexudad
ba-kalmar,
'not the sharpest pencil in the case'
- (16) βγάλω από τα ρούχα μου (EL) – vjaino apo ta rucha mu
'take-off of the clothes my'
- (17) nincs ki a négy kereke (HU)
'he does not have all the four wheels'
- (18) estar de saco cheio (PT)
'to have one's bag full'
- (19) كَرْدَن كَفِ كَسِي دَهَن (FA) – dahānē kasy kaf kardan
'mouth someone's to foam'

At the “Multiword games” session of the 7th Parseme general meeting all participants (except organizers) were divided into four groups and they worked together to find the solutions. There were between 8 and 13 participants in each group. The time was limited to 15 minutes. All groups performed extremely well: the Germanic group solved all 12 problems, the Romance group solved 10, the Other languages group 9, and the Slavic group 8.

The games for all four groups as well as the spreadsheet with all users' input for 10 senses can be found at the event's page.⁶

5 MWE-based pantomime

For this game we used only English MWEs so that everybody could participate, presuming that everybody in the audience knew English. Most of English MWE examples were selected from famous seminal papers on MWEs (Baldwin and Kim, 2010; Sag et al., 2002), which all PARSEME COST action participants must have read. A few other were chosen either as being generally known, e.g. *all of a sudden*, or as being mentioned many times during PARSEME meetings, e.g. *it rains cats and dogs*. All chosen MWEs were ranked as being easy, medium or difficult according to two criteria: (i) is the MWE well known? (ii) is it easy to demonstrate in a pantomime game (e.g. adverbs are harder to demonstrate than nominal or verbal MWEs)?

English native speakers did not participate in this game; they instead had the role of referees. The audience was divided into two groups. Each group received a set of English MWEs: 4 easy, 4 medium and 1 difficult

⁶ Games are at <https://goo.gl/xM9irG>

MWE. One person of a group showed the MWE by pantomime, while the rest of the group guessed. Guessing time was limited to 1, 2 and 3 minutes for the easy, medium and hard MWEs, respectively. The demonstrator was allowed to point by fingers the number of constituents the MWE consisted of, otherwise she/he could use all non-verbal means of communication to try to explain what a particular constituent was.

As an example, MWEs of one group were:

- Easy MWEs (1 minute):
 1. *pain in the neck* A SOURCE OF ANNOYANCE, A NUISANCE;
 2. *take the bull by the horns* TO CONFRONT A PROBLEM HEAD-ON AND DEAL WITH IT OPENLY;
 3. *spill the beans* TO GIVE AWAY A SECRET OR A SURPRISE;
 4. *strike while the iron is hot* WHEN YOU HAVE AN OPPORTUNITY TO DO SOMETHING, DO IT BEFORE YOU LOSE YOUR CHANCE.
- Medium MWEs (2 minutes):
 5. *cranberry word* – in linguistic terminology, a cranberry morpheme (or fossilized term) is a type of bound morpheme that occurs only in MWEs and cannot be assigned an independent meaning or grammatical function, as in *to go astray*;
 6. *red tape* THE COLLECTION OR SEQUENCE OF FORMS AND PROCEDURES REQUIRED TO GAIN BUREAUCRATIC APPROVAL FOR SOMETHING, ESPECIALLY WHEN OPPRESSIVELY COMPLEX AND TIME-CONSUMING;
 7. *sell like hot cakes* TO BE SOLD VERY FAST;
 8. *trip the light fantastic* TO DANCE, IN PARTICULAR ENGAGE IN BALLROOM DANCING.
- Difficult MWE (3 minutes):
 9. *by and large* ON THE WHOLE, EVERYTHING CONSIDERED

At the end, all MWEs were guessed by the audience although we have to admit that referees were not very strict over the time spent. However, one MWE was guessed before the demonstration even started. It was enough that the demonstrator said “It is an MWE that everybody knows” and the solution was offered immediately: *kick the bucket* TO DIE.

The MWEs to guess for both groups can be found at the event’s page.⁷

⁷ Available at <https://goo.gl/QVVUgC>

6 Quiz with nationalities

As mentioned in Section 3, part of the collected MWEs refer to nationalities and mostly convey positive or negative stereotypes about them. We gathered 80 such expressions from 23 countries about 32 various nationalities or groups of nationalities in total. These data were then manually transformed into a graph-based representation, overlying a political map of Europe, so that an arc goes from country A to country B if the language spoken in A uses an MWE referring to the nationality of B. This representation oversimplifies the country-to-nationality correspondence but is an acceptable approximation for our aim. The color of each arc depends on the polarity of the stereotype expressed by the MWE:

- red: negative stereotype,
- orange: neutral,
- green: positive stereotype.⁸

For instance, as shown in Fig. 1, a red arc going from Latvia to Sweden represents expression (20), an orange arc from Greece to France stands for (21), and a green arc from Malta to the UK corresponds to (22). The arrows are larger or narrower, depending on the number of the MWEs having the same source and target countries, as well as polarity.

- (20) zviedru gardīnes (LV)
‘Swedish curtains’ \Rightarrow PRISON
- (21) γαλλικό κλειδί (EL) – jaliko klidi
‘French key’ \Rightarrow ADJUSTABLE WRENCH
- (22) għandu ħin Ingliż (MT)
‘to have an English timing’ \Rightarrow TO BE PUNCTUAL

This representation facilitated the analysis of the data set. For instance, we had expected a large number of MWEs to concern neighboring countries, due to historical reasons. Such expressions were however not very frequent in our data set – see examples (23)–(28) and Figure 2.

- (23) leben wie Gott in Frankreich (DE)
‘live like God in France’ \Rightarrow TO LIVE IN LUXURY

⁸ Colors cannot be seen in the printed version of this paper, but are visible in the on-line version.

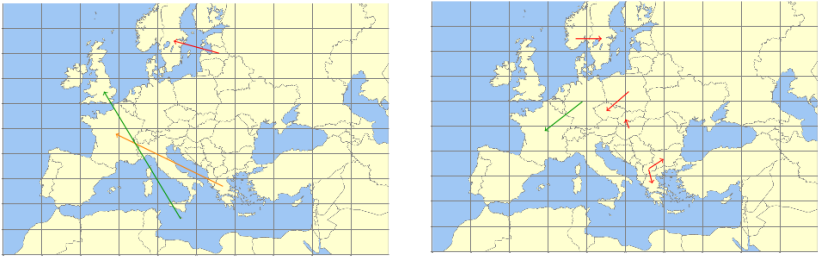


Figure 1. Graph-based representation **Figure 2.** MWEs from examples (23)–(28) concerning neighboring countries, referring to nationalities and their polarities.

- (24) tótágast áll (HU)
‘it stands as a Slovakian hanger’ \Rightarrow IT IS UPSIDE DOWN
- (25) бугарски чорапи (MK) – bugarski chorapi
‘bulgarian socks’ \Rightarrow BEARFOOT
- (26) солунски диреци (MK) – solunski direci
‘Thessaloniki pillars’ \Rightarrow FAT (USUALLY FEMALE) LEGS
- (27) svenskemetoden (NO)
‘the Swedish method’ \Rightarrow AN EASY SOLUTION
- (28) czeski film (PL)
‘Czech film’ \Rightarrow A SITUATION IN WHICH NO ONE UNDERSTANDS ANYTHING

Another interesting phenomenon is the one of minimal non-empty cycles, i.e. a situation where several, usually two, countries use negative-polarity MWEs about each other. We found two instances of such cycles, shown in examples (29)–(32) and Figure 3.

- (29) filer à l’anglaise (FR)
‘to leave in the English style’ \Rightarrow TO LEAVE WITHOUT SAYING GOOD-BYE
- (30) to take a French leave (EN)
TO LEAVE WITHOUT SAYING GOODBYE
- (31) saoul comme un polonais (FR)
‘as drunk as a Pole’ \Rightarrow VERY DRUNK



Figure 3. MWEs from examples (29)–(32), conveying reciprocal negative (28) using antonymy to express stereotypes.



Figure 4. MWEs from examples (23)–(32), conveying reciprocal negative (28) using antonymy to express sarcasm.

- (32) francuski piesek (PL)
 ‘a *French doggy*’ \Rightarrow SOMEONE CAPRICIOUS AND DIFFICULT TO PLEASE

As far as linguistic phenomena exploited in the stereotype-driven MWEs are concerned, antonymy was attested as expression of sarcasm. The relevant examples (33)–(35) are illustrated in Figure 4.

- (33) наивна к’о француска собарица (SR) – naivna k’o francuska sobarica
 ‘as *naïve/inexperienced as a French maid*’ \Rightarrow VERY EXPERIENCED BUT BEHAVING AS IF IT WAS THE OPPOSITE
- (34) cald nemțesc (RO)
 ‘*German warmth*’ \Rightarrow COLD
- (35) a merge drept ca neamțul (RO)
 ‘to walk as *straight as a German*’ \Rightarrow TO STAGGER LIKE A DRUNK MAN

A phenomenon similar to those exploited in the game on general meanings was also observed in the MWEs with nationalities: the same meaning was expressed in different languages by referring to different nationalities, as in examples (29)–(30). Two popular meanings of this kind were TO PRETEND NOT TO UNDERSTAND and TO BE DRUNK/TO DRINK A LOT. They were represented in our data by 4 and 3 different nationalities, respectively, as shown in examples (36)–(39), (31), (40)–(42) and Figures 5–6.⁹

⁹ Note that when a source or a target country is outside Europe, e.g. China in example (36), the arrow has its starting or ending point outside the map.

- (36) κάνω τον Κινέζο (EL) – kano ton Kinezo
‘to make the Chinese’
- (37) udać Greka (PL)
‘to pretend being a Greek’
- (38) hacerse el sueco (ES)
‘to act like a Swedish’
- (39) правити се Енглеz (SR) – praviti se Englez
‘to pretend being an Englishman’
PRETEND NOT TO UNDERSTAND
- (40) pít jako Dán (CZ)
‘to drink like a Danish’
- (41) пие како Рус (MK) – pie kako Rus
‘to drink like a Russian’
- (42) пиян като казак (BG) – piyan kato kazak
‘as drunk as a Cossack’
TO BE DRUNK/TO DRINK A LOT



Figure 5. MWEs from examples (36)–(39) expressing the meaning TO and (40)–(42), expressing the meaning PRETEND NOT TO UNDERSTAND. **Figure 6.** MWEs from examples (31) TO BE DRUNK/TO DRINK A LOT with different nationalities.

Certain MWEs expressed stereotypes concerning not only selected nationalities but larger populations. Notable examples stem from Swahili and concern the white people – cf. examples (43)–(44).

- (43) Mzungu wa reli (SW)
‘the railway white person’ ⇒ IGNORANT

- (44) Mzungu wa unga (SW)
 ‘white man of flour’ \Rightarrow DRUG DEALER

Some results of the data analysis were offered to the audience as simple quiz questions. For instance, three European countries gathered ex aequo the highest number of the incoming edges, and the audience was to guess which countries were concerned. A similar quiz concerned the country targeted by MWEs from several languages of a given geo-political region, namely the Balkans. The examples relevant to both quizzes and the respective graphs are shown in Appendix 2.

The graph representing all collected MWEs referring to nationalities is shown in Figure 7. Out of the total number of 80 MWEs it represents, only 7 refer to positive stereotypes, 3 are neutral and the remaining 70 expressions convey negative stereotypes. The whole dataset is freely available at <https://goo.gl/U2kKeR>.



Figure 7. European graph of MWEs referring to nationalities.

7 Additional outcome

Many of the collected MWEs were not exploited for the MWE games but are a valuable source of observations. Those include above 100 examples provided by the contributors as particularly opaque or funny.¹⁰ They can be analyzed in terms of figuration and transparency (Sheinflux et al., 2017). Figuration refers to the degree to which the MWE can be assigned a literal meaning. Most of our examples in the opaque/funny category are figurative and part of the amusing effect stems from the unexpectedness of the image conveyed by their literal reading, as in examples (45)–(48).

- (45) viajar na maionese (PT)
‘to travel in the mayonese’ ⇒ TO SAY/THINK SOMETHING SILLY/CRAZY
- (46) vzít nohy na ramena (CZ)
‘to take one’s legs onto one’s shoulders’ ⇒ TO RUN AWAY
- (47) prendre des vessies pour des lanternes (FR)
‘take bladders for lanterns’ ⇒ TO BE WRONG
- (48) puust piiliites (LV)
‘to blow ducks’ ⇒ TO TELL FAIRYTAILS

Transparency refers to how easy it is to recover the motivation for the idiom due to the relationship between its literal and idiomatic reading. The amusing effect can originate from a high transparency, due to hyperbole or sarcasm, as in examples (49)–(52) or, conversely, from non-transparency, as in (53)–(57).

- (49) der er ingen ko på isen (DA)
‘there is no cow on the ice’ ⇒ THERE IS NOTHING TO WORRY ABOUT
- (50) κόβω καρφιά (EL) – kovo karfia
‘to cut nails’ ⇒ TO FEEL VERY COLD
- (51) itatja az egereket (HU)
‘he makes the mice drink’ ⇒ HE IS CRYING
- (52) rope på elgen (NO)
‘call on the moose’ ⇒ TO VOMIT
- (53) вешать лапшу на уши (RU) – veshat’ lapshu na ushi
‘to hang noodles on somebody’s ears’ ⇒ TO DECEIVE, TO TELL LIES
- (54) си легна на брашното (MK) – si legna na brashnoto
‘to lie down on the flour’ ⇒ TO RESIGN ONESELF TO ONE’S FATE

¹⁰ The whole dataset is available from the event’s page: <https://goo.gl/6R6ZaD>

- (55) tirar los tejos a alguien (ES)
'to throw tiles to someone' ⇒ TO FLIRT
- (56) i jomarkə joti (LT)
'to ride to the marketplace' ⇒ TO BE BLUNT (ABOUT KNIVES OR SCISSORS)
- (57) kuzunguka mbuyu (SW)
'to go round a baobab tree' ⇒ TO BRIBE

The analysis of data gathered from users on MWEs for 10 general meanings (cf. Section 4) also showed that many same or similar metaphors are used in different languages for the same meaning. We selected all those that were shared by at least two languages and offered a new table to the community to fill in the similar MWEs, if they exist, in their languages. It turned out that the most frequent was the rhetoric figure simile *to work like a...* where the empty slot can be filled with *dog* (EN, FR, EL, HR, SR), *horse* (ET, HR, LV, SR), *donkey* (HE, MK), *mule*, *cattle*, and *worm* (HR, SR), and *bullock* (PL). Comparisons with nationalities and ethnic groups included *a Negro* (HR, MK, SR), *a Turk* (LV), and *a Chinese* (ES).

For the meaning TO DO SOMETHING USELESS a number of MWEs were collected:

- (58) מִים תַּחַן (HE) – Taxan mayim,
'grind water'
- (59) толочь воду в ступе (RU) – toloch' vodu v stupe
'to grind water in a mortar'
- (60) a bate apa în piuă (RO)
'to beat the water in the mill'
- (61) pestare l'acqua nel mortaio (IT)
'to grind water in a mortar'
- (62) pretakati iz šupljeg u prazno (HR)
'to pour from hollow into empty'
- (63) کُـبـِـدَن هَـفَن دَر آب (FA) – āb dar hāvan koobidan,
'pounding water in a pounder'
TO DO SOMETHING USELESS

The table of similar MWEs across languages for 10 general meanings is publicly available at <https://goo.gl/hf6Wup>.

8 Conclusions and perspectives

We have presented an effort towards constructing a set of games based on collectively gathered examples of multiword expressions in many languages. These games, aimed at community building, proved successful, if judged by the enthusiastic participation and positive feedback from the network members. The major reason of this success is most probably the very nature of MWEs, whose semantic non-compositionality can be easily exploited for its surprising and amusing effect.

These games were addressed to experts in linguistics and natural language processing, and in MWEs in particular. Extending them to a larger public would be worth examining, notably in language acquisition, foreign language learning, journalism, international relations and many others. We also believe that by extending the scope and coverage of this study to new expressions, languages and countries, interesting correlations between historical and linguistic phenomena might be observed.

Abbreviations

BG ‘Bulgarian’, CZ ‘Czech’, DA ‘Danish’, DE ‘German’, EL ‘Greek’, EN ‘English’, ES ‘Spanish’, EST ‘Extrasensory perception’, FA ‘Farsi’, FR ‘French’, GWAP ‘game with a purpose’, HE ‘Hebrew’, HR ‘Croatian’, HPU ‘human processing unit’, HU ‘Hungarian’, IT ‘Italian’, LT, ‘Lithuanian’, LV ‘Latvian’, MK ‘Macedonian’, MT ‘Maltese’, MWE ‘multiword expression’, NLP ‘natural language processing’, NO ‘Norwegian’, PL ‘Polish’, PT ‘Brazilian Portuguese’, RO ‘Romanian’, RU ‘Russian’, SR ‘Serbian’, SW ‘Swahili’

1 Solutions for the General meaning game of the Slavic group from Section 4

- (8) – a (TO DIE)
- (9) – b (TO BE TIRED)
- (10) – b (TO BE TIRED)
- (11) – a (TO DIE)
- (12) – d (TO BE ANGRY)
- (13) – a (TO DIE)
- (14) – c (TO BE STUPID)
- (15) – c (TO BE STUPID)

- (16) – d (TO BE ANGRY)
- (17) – c (TO BE STUPID)
- (18) – d (TO BE ANGRY)
- (19) – b (TO BE TIRED)

2 Solution to the quizzes on the (globally and locally) highest number of incoming edges from Section 6



Figure 8. France, Russia and UK: the country gathering the highest number of incoming edges. Sample MWEs are listed in Ex. (64)-(66)



Figure 9. Turkey: the country gathering the highest number of incoming edges stemming from the Balkan region. Sample MWEs are listed in Ex. (67)-(70)

- (64) (só) pra inglês ver (PT)
'(just) for English people to see' ⇒ (JUST) FOR THE SAKE OF APPEARANCES
- (65) Francuski poljubac (HR)
FRENCH KISS
- (66) en by i Rusland (DA)
'a town in Russia' ⇒ SOMETHING ENTIRELY IRRELEVANT
- (67) Puši k'o Turčin (HR)
a fuma ca un turc (RO)
'to smoke like a Turk' ⇒ TO SMOKE A LOT
- (68) γίνουμαι Τούρκος (EL) – jinome Turkos
'to become a Turk' ⇒ TO BECOME FURIOUS

- (69) turcul plătește (RO)
 ‘the Turk pays’ \Rightarrow ONE HAS TO TAKE RESPONSIBILITY FOR SOMETHING THAT ONE DIDN’T DO
- (70) пролазити као поред турског гробља (SR) – prolaziti kao pored turskog groblja
 ‘to go by as near a Turkish cemetery’ \Rightarrow TO PASS BY WITHOUT PAYING ATTENTION

References

- Von Ahn, Luis and Laura Dabbish. “Designing Games with a Purpose”. *Commun. ACM* Vol. 51, no. 8 (2008): 58–67. <http://doi.acm.org/10.1145/1378704.1378719>
- Baldwin, Timothy and Su Nam Kim. “Multiword Expressions.”. *Handbook of natural language processing* Vol. 2 (2010): 267–292.
- Cooper, Seth, Adrien Treuille, Janos Barbero, Andrew Leaver-Fay, Kathleen Tuite et al.. “The Challenge of Designing Scientific Discovery Games”. In *Proceedings of the Fifth International Conference on the Foundations of Digital Games*, FDG ’10, New York, NY, USA: ACM, 40–47, 2010. <http://doi.acm.org/10.1145/1822348.1822354>
- Gurevych, Iryna and Torsten Zesch. “Collective Intelligence and Language Resources: Introduction to the Special Issue on Collaboratively Constructed Language Resources”. *Language Resources and Evaluation* Vol. 47, no. 1 (2013): 1–7.
- Howe, Jeff. “The rise of crowdsourcing”. *Wired magazine* Vol. 14, no. 6 (2006): 1–4.
- Kittur, Aniket, Jeffrey V. Nickerson, Michael Bernstein, Elizabeth Gerber, Aaron Shaw et al.. “The Future of Crowd Work”. In *Proceedings of the 2013 Conference on Computer Supported Cooperative Work, CSCW ’13*, New York, NY, USA: ACM, 1301–1318, 2013. <http://doi.acm.org/10.1145/2441776.2441923>
- Lafourcade, Mathieu, Alain Joubert and Nathalie Le Brun. *Games with a Purpose (GWAPS)*. John Wiley & Sons, 2015.
- Millour, Alice and Karën Fort. “Why do we Need Games? Analysis of the Participation on a Crowdsourcing Annotation Platform”. In *Games4NLP*, 2017.
- Mitrović, Jelena. “Crowdsourcing and its application”. *INFOtheca* Vol. 14, no. 1 (2013): 37–46.

- Sag, Ivan, Timothy Baldwin, Francis Bond, Ann Copestake and Dan Flickinger. “Multiword Expressions: A Pain in the Neck for NLP”. *Computational Linguistics and Intelligent Text Processing* (2002): 189–206.
- Sheinflux, Livnat Herzig, Tali Arad Greshler, Nurit Melnik and Shuly Wintner. *Representation and Parsing of Multiword Expressions*, ed. Yannick Parmentier and Jakub Waszczuk, 5–38. Chapter Verbal MWEs: Idiomaticity and flexibility. Berlin: Language Science Press, 2017.
- Tuite, Kathleen. “GWAPs: Games with a Problem.”. In *FDG*, 2014.

The Use of the Omeka Platform for Digital Libraries in the Field of Mining

UDC 004.738.5:027.7]:004.42OMEKA

DOI 10.18485/infotheca.2017.17.2.2

Aleksandra Tomašević
aleksandra.tomasevic@rgf.bg.ac.rs

Biljana Lazić
biljana.lazic@rgf.bg.ac.rs

Dalibor Vorkapić
dalibor.vorkapic@rgf.bg.ac.rs

Mihailo Škorić
mihailo.skoric@rgf.bg.ac.rs

Ljiljana Kolonja
ljiljana.kolonja@rgf.bg.ac.rs

*University of Belgrade
Faculty of Mining and Geology*

Translated by: Marko Vitas

ABSTRACT: This paper will introduce Omeka, a platform for presentation of digital collections and a system for the management of their content. We will illustrate its application in the field of technical sciences (more specifically, in the field of mining) on the example of the digital library ROmeka@RGF. We have decided to use Omeka because it is simple, because it possesses comprehensive supporting documentation and because it does not require any attainment in information sciences, which makes it accessible for most users, and especially for mining engineers, to whom this digital library is chiefly intended. Documents assembled and stored in this digital library will serve as a basis for future research, extraction of terminology, tagging, extraction of knowledge etc.

KEYWORDS: Omeka, digital library, mining.

PAPER SUBMITTED: 2 September 2017

PAPER ACCEPTED: 11 October 2017

1 Introduction

For the purposes of this paper, a digital library ROmeka@RGF¹ was created in order for the academic texts in the field of mining to be assembled, systematized, processed and stored. This is meant to serve as an important basis not only for various linguistic and terminological research studies, but also for a whole array of tasks related to the knowledge engineering (not

¹ <http://romea.rgf.rs>

only the extraction of knowledge). Documents have been assembled, processed and stored in the digital library in the course of the projects financed by the Ministry of Education, Science and Technological Development of the Republic of Serbia. Inspiration for this project came from the need to make the texts in the field of mining more readily accessible to both mining engineers and linguists. In the course of the preliminary research for this project no digital library for the field of mining was discovered, except for the journal „Underground Mining Engineering” (original title in Serbian: „Podzemni radovi”), which is a component of the bilingual library „Bibliša”². Therefore, the digital library ROmeka@RGF can be considered the first of its kind in Serbia. According to our knowledge, the only example of a digital library on the Omeka web platform, in the field of geology (a field narrowly connected to mining) is the digital library „Geological books of Philippe Glangeaud” (*Carnets géologiques de Philippe Glangeaud*³). This digital library had been developed on the Clermont university.

Section 2 of this paper contains a description of the developmental environment of the web platform Omeka and a review of the software’s most useful plugins. Section 3 contains a description of the most important elements of the digital library and the ways these elements are created. Section 4 contains a description of the digital library ROmeka@RGF itself. Section 5 explains how to search through the stored textual resources with help of tools and resources available for Serbian. Section 6 contains possible implementations of the TEI guidelines⁴. Section 7 contains conclusions and ideas for future research.

2 The Web Platform Omeka

Omeka, the web platform for displaying digital collections and the system for managing their contents⁵, has been used for the creation of the digital library ROmeka@RGF. Omeka was developed in the Roy Rosenzweig Center for History and New Media at the George Mason university in Virginia⁶. It belongs to the Open Source Software group and has the General Public Li-

² <http://jerteh.rs/biblisha/ListuDokumenata.aspx?JCID=2&lng=en>

³ <http://bibliotheque.clermont-universite.fr/glangeaud/>

⁴ TEI: Text Encoding Initiative <http://www.tei-c.org/index.xml>

⁵ Content Management System (CMS)

⁶ Roy Rosenzweig Center for History and New Media (RRCHNM), <https://rrchnm.org/>

cense (GPL v3.0)⁷. This means that its source code is publicly available and that each user may upgrade or adapt it to meet his or her own needs. While the platform’s original target audience were academic institutions studying cultural heritage, today it is used by researchers in many different fields.

Omeka’s flexibility and user-friendliness earned it a top position among the software of its kind. Its main features are: attractive and flexible visual design, simple installation, capacity to spread which enables replacement of the existing functions and the addition of the new ones, flexible approach to meta-data, support for web standards (CSS, HTML, RSS), import and export of the data in the standardized formats (RDF, CSV, XML, JSON) (Kucsma et al., 2010).

Since the software had not been designed for IT-experts and does not require a high level of proficiency in information sciences, it allows the users to focus on the contents of the digital library, as well as on the description and interpretation of its contents, rather than on programming. Furthermore, the platform disposes of functions for catalogization and presentation of digital items. These functions are based on the Dublin Core, which ensures that the description and organization of digital items is standardized.

2.1 The developmental environment of Omeka

There are two basic versions of the web platform Omeka:

- Omeka.neta, a version which does not require its own server. Its file storage capacity is limited to 500MB, the size of a single file is limited to 64MB and the number of available plugins is limited to fifteen. It does not allow for functionality adjustments and options for the platform’s appearance adaptation are severely curbed. The lite version can be upgraded, but this requires the user to pay between 35 and 1000 dollars a year, depending on the package.
- Omeka.org, a complete version which can be installed on a local disk or as a virtual machine and allows all functionality adjustments.

Certain adjustments are needed on the server before Omeka can be installed and they include the installment of: the web (HTTP) server Apache, the database management system MySQL (either 5.0 or a newer version) and the program language interpreter PHP (either 5.3.2 or a newer version). The following distributions of the operating system Linux allow Omeka to work

⁷ <https://www.gnu.org/licenses/gpl-3.0.en.html>

steadily: Fedora, OpenSuse and Ubuntu. ROmeka@RGF has been installed on a virtual machine with the operating system Ubuntu 15.10.

The installation of the web platform is initiated by creation of the MySQL base with administrator privileges. The newest version of Omeka is downloaded from the platform's official web page and unpacked. The resulting folder (hereinafter: *omeka-root*) is stored either in the root folder of the web server or in one of its subfolders. Certain changes have to be made in the file *omeka-root/db.ini* (the following field values: database host, username, password and database name) in order for the database management system MySQL to become operational. All folders containing Omeka are required to have a permit. After the installation had been completed it is necessary to adjust privileges. The installation is launched through the web-reader by entering the IP address or domain of the digital library, where the administrator account and the site name have to be defined.

The web platform is multilingual and translated, either fully or partially, into 50 languages. The Serbian version is among the ten that have been fully translated.

2.2 The Omeka software plugins

The appearance of the web platform can be altered and adapted: a user is free to either choose among fifteen offered themes or create his or her own theme. Meanwhile, functionality is increased through plugins. Ninety plugins have been developed and adapted to different versions of Omeka. The users of Omeka have developed over 300 plugins of their own, but since they were mainly designed for the older versions of Omeka, further adjustment is needed for the newest version of the platform. Depending on their purpose, plugins can be classified into several categories: plugins for mass creation of collections and items, plugins for content management, plugins for file inspection, plugins for the community contribution descriptions, plugins for geospatial processing and map navigation. Here we will review several plugins, both such that were used for the creation of our digital library and such that we deemed important and useful for all and any users of Omeka.

Archive Repertory enables the user to keep the imported files on the server with their source names. It also allows files to be gathered in a hierarchical structure, based either on items or on collections containing these items. The ability to keep source names makes it easier to read URL-addresses and to manage files.

Bulk Metadata Editor enables the user to search and update the meta-data for a large number of items simultaneously and in a fast and simple way. The first step is to make a selection of items based on various criteria (for instance, a selection can encompass: all items of a digital library, items located in a certain collection and items incorporating meta-data which meet one or several criteria). The second step requires the user to choose the piece of meta-data which should be altered. In the third and concluding step, the meta-data is finally altered (for instance, one can alter the following: search and replacement of text, addition of new meta-data in the selected field, addition of text to the existing meta-data, removal of duplicates and empty fields in the selected description of an item, removal of duplicated files in the selected items or removal of all existing meta-data in the selected fields).

Catalog Search makes it possible to search through other catalogues by using the field dc: subject. These catalogues include: Archive Grid, Digital Public Library of America, Google Books, Google Scholar, Hathi Trust, JS-TOR, Library of Congress, WorldCat. It is equally possible to add links to the catalogues of other institutions.

*COinS*⁸ arranges for the meta-data about citations for every item to be embedded into web-pages of the digital libraries. Once activated, *COinS* provides visibility for objects on on-line platforms such as Zotero⁹ by automatically embedding meta-data about citations into other web-pages. Additionally, individual items of any Omeka site can be added to the library of the platform Zotero, while the simultaneous addition of several items can be implemented through the scripts operating in the background. *COinS* plug-in makes the research and the interoperability with other systems much easier.

Collection Tree secures a visual representation of the hierarchical structure in the digital library, consequently making the browsing through them easier.

CSV Import enables mass import of meta-data, labels and files, which are represented by a table in the CSV format. If the titles of the columns are in accordance with the Dublin Core, then the data are mapped automatically; if they are not, then it is necessary to map the data manually.

Drop Box makes makes the import of data located on the server (in the catalogue */plugins/Dropbox/files*) simpler for the administrator. The admin-

⁸ ContextObjects in Spans (COinS), http://omeka.org/codex/Plugins/Coins_2.0, accessed May 28, 2017.

⁹ <https://www.zotero.org/>

istrative interface allows the data to be imported either individually or on a large scale through selection from a list.

Dublin Core Extended expands the list of meta-data of the Dublin Core. This secures a total annotation of items. To 15 original elements (title, creator, subject, description, publisher, contributor, date, type, format, identifier, source, language, relation, coverage, rights) additional 40 are added: abstract, access rights, accrual method, accrual periodicity, accrual policy, alternative title, audience, date available, bibliographic citation, conforms to, date created, date accepted, date copyrighted, date submitted, audience education level, extent, has format, has part, has version, instructional method, is format of, is part of, is referenced by, is replaced by, is required by, date issued, is version of, license, mediator, medium, date modified, provenance, references, replaces, requires, rights holder, spatial coverage, table of contents, temporal coverage, date valid.

Geolocation arranges for the information about locations relative to the digital items to be added to the maps and enables the user to search through them.

Item Relations enables the creation of the relation between the digital objects. This plug-in follows the RDF model for defining relations between items. This model is represented as an RDF graph made out of RDF triplets: subject-predicate-object. For example, if one digital item (RDF subject) is a part of another digital item (RDF object), then a correlation *isPartOf* (RDF predicate) is established between them. Similarly, if one digital item is a version of a document, then between these digital items a correlation *isVersionOf* is established. This is how the RDF triplets are formed. These triplets enable the text to be searched through at a later point by using the techniques of the semantic web.

Hide Elements enables the user to choose meta-data which will be hidden on the import form, on the web-page of the administrator and/or on the publicly available web-page, and also on the form for the browsing of the meta-data.

*METS Export*¹⁰ enables the export of digital items such as METS XML files, individual files, collections or whole digital libraries. It is supported by the Initiative of the Digital Library Federation¹¹, which suggests an XML scheme of meta-data for the management of items in a digital library and

¹⁰ The Metadata Encoding and Transmission Standard,
<https://www.loc.gov/standards/mets/METSOverview.v2.html>

¹¹ Digital Library Federation, <https://www.diglib.org/?s=mets>

for their exchange between repositories or between a repository and a user. It is most notably used for gathering and keeping of documents contained in a digital item, given their number and variety. *METS Export* links several digital documents and enables navigation between them. Additionally, it contains technical information necessary for managing digital items: formats, technological characteristics, ways of scanning, digital transformations. *METS Export* does not require a specific group of meta-data to be entered for a digital document, but instead allows the creator of the meta-data to decide which meta-data he or she would like to enter for that purpose. Descriptive meta-data for METS can easily be downloaded from the Dublin Core (Тртовац, 2016).

Neatline, *NeatlineFeatures*, *NeatlineSmile*, *NeatlineText*, *NeatlineTime* and *NeatlineWaypoints* represent a series of plug-ins which allow the spatial and temporal points on the items map in the digital library to be interconnected. It also allows the documents to be connected with the *Neatline* exhibition. These plug-ins have not been activated in the digital library ROMeka@RGF because they are incompatible with the plug-in *Geolocation*, which is far more useful for engineers.

OAI-PMH Harvester gathers meta-data from an *OAI-PMH*¹² data supplier, maps them in a local database and imports them. It can be used both on a short-term and on a permanent basis for updating, synchronization and distribution of the systems. At the present time, the formats it is able to import are the Dublin Core and *METS*.

OAI-PMH Repository prepares items for exchange and is functionally in an inverse relationship to the plug-in discussed previously (*OAI-PMH Harvester*). It supports the Dublin Core, MODS and *METS*.

PDF Text enables the optical recognition of the text characters¹³, the extraction of the text from the PDF format and the search through the text. If the text is not read satisfactorily it is possible to correlate it or to import a new textual document to the provided field.

Reference adds pages equipped with alphabetical index of elements, which, defined in advance, in turn allow browsing the meta-data which were also set in advance.

Search By Metadata allows the administrator to define metadata for advanced searches on a HTML page.

¹² Open Archives Initiative Protocol for Metadata Harvesting,
<https://www.openarchives.org/pmh/>

¹³ Optical character recognition (OCR)

SimplePages enables the administrator to create dynamical PHP pages without requiring a specific informatic attainment.

SimpleVocab and *SimpleVocabPlus* enable the creation of controlled dictionaries and their synchronization on a cloud. In the digital library ROMEKA@RGF a controlled authorial dictionary had been created. This secured a consistent export and made searching easier.

TEI Display renders a TEI file (which had been prepared and joined to an item) into a visually clear shape. The understood XSLT transformation enables two presentation modes: either the whole document or its individual units are presented. While the first mode presupposes the transformation of the whole document into a HTML, the second presents the content of the document (div1 or div2). This latter option is particularly suited for larger documents. Both the presentation modes and XSLT transformations can be adapted, while the metadata from the TEI heading can be automatically mapped in fields of the Dublin Core for items and files.

3 The creation of the digital library

According to one of the most frequently cited definitions of digital libraries, conceived by William Y. Arms, digital libraries are controlled and systematically organized collections of information, with adjoined services, which are stored in digital format and accessible through the web. The common feature of all digital libraries is the fact that information is organized on computers and accessible via internet, together with procedures for their selections, organization (so that they can become more readily accessible) and archiving (Arms, 2000).

Digital libraries are collections of digital items which are stored on the web in form of various digital data (e.g. text, image, sound, video, animation) or in form of combinations of digital items (multimedia). They are described by various metadata and connected to other informational services. End users can access and use them without any temporal or spatial limitations. They can also create new digital items and update the old ones (Трговац, 2016).

Basic elements of a digital library in Omeka are:

- items,
- collections and
- web pages.

The user is free to import an unlimited number of items, documents, signs and collections. The only limitation is that one item can be adjoined to one collection only (Figure 1)¹⁴. Each of the aforementioned elements of the digital library is liable to a complete visibility control on the web. This control encompasses everything from individual metadata to all of the element in the whole.

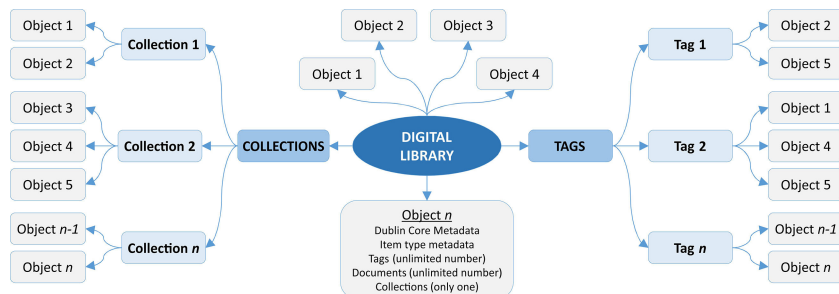


Figure 1. A diagram of the elements of a digital library in Omeka

3.1 Items

The Omeka web platform has been designed for presentation of objects, as they are the fundamental element of every digital library. Therefore, the creation of a digital library necessarily begins by the creation of items.

Depending on their type, items can vary (and the same goes for units, archives, sources or resources). User is provided with a list of 15 basic types of items. There is also a possibility to add new ones if needed. The following item types are accessible:

- *Moving Image* – video recordings of all sorts: animations, films, television shows;
- *Sound* – audio recordings of all sorts: audio compact disks, recorded speech or sound;
- *Oral History* – information obtained through interviews with persons in possession of first-hand knowledge;

¹⁴ https://omeka.org/codex/Managing_Items, access date 19. May 2017

- *Still Image* – visual presentation of texts, images, drawings, graphic design, plans or maps;
- *Website* – HTML pages with adjoined images, audio and video files etc.;
- *Event* – temporally limited phenomena, for example: an exposition, a web conference, a workshop, a tea party, a fire, a battle, a trial, a wedding;
- *Email* – textual messages with optional attachment(s), sent by one person to the other person or persons;
- *Lesson Plan* – a detailed description of teaching process throughout a course;
- *Person*;
- *Interactive Resources* – web pages, multi-media classroom items, chat services;
- *Dataset* – coded data in a defined structure: lists, tables and databases;
- *Physical Objects* – three-dimensional solid inanimate objects, represented in digital libraries by types such as moving or static pictures et al.
- *Services* – for example: copying services, banking services, interlibrary loans or web servers;
- *Software*;
- *Hyperlinks* – a link or a reference to a different resource on the Internet.

Items are collections of:

- metadata from the Dublin Core describing the digital item itself,
- item type metadata,
- tags,
- documents.

Metadata are structured in such a way that enables them to describe, explain, identify, locate or enables them to make retrieval, use and managing of the information source easier in some other way (Hodge, 2001). They can be:

- *descriptive* – they describe resources that are needed in order to find and identify the source of information. They contain some basic elements such as: title, author, publisher, place, year, language, unique identifier, description, keywords, subject headings, abstract etc.;
- *structural* – they describe the structure of complex resources: types, versions, connections between digital items and other features, connections between the original document and its versions, including the data about changes and other features;

- *administrative* – they offer information on use and managing of resources in connection to the intellectual right. They can be:
 - metadata on copyright (which define the management of rights to access a digital item in accordance with the authorial rights and with the protection of the intellectual property);
 - technical metadata (which contain data on the creation date, technical details of the source, size and type of the file, access to the source, data on all the changes and format of the presentation);
 - metadata on conservation of a digital item;
 - metadata on use (they refer to active tracking of numbers of users who visit and use a particular content, as well as to tracking of use of a digital item content in a new context or in a new version by downloading the metadata and the digital item for a different digital library).

First, items are described by the metadata from the expanded Dublin core. The type of item is also described by the metadata, which are different depending on the type. For example, if the item is „*oral tradition*”, then the metadata describing it are: the person conducting the interview, the person being interviewed, the location, the transcript, the duration of the interview etc. If the item is „*web page*”, then only URL is entered, while for the item „*person*” corresponding metadata are: the date of birth, the place of birth, the date of death, profession, biography, bibliography etc. It is possible to describe an item through the metadata that will make it more visible in the scope of the Zotero platform (Figure 2).

Documents which are attached to the items can be imported either individually or on a large scale, either by the use of the DropBox plugin (Figure 3) or by the import from a file in the CSV format. Documents can be found in different formats, and some of the commonest are:

- for text: txt, css, csv, rtf, rtx, doc, docx, pdf, pps, ppt, pptx;
- for tables: xls,xlsx;
- for databases: mdb;
- for images: bmp, gif, jpeg, jpg, tiff, png;
- for video recordings: avi, divx, mpeg, mov, mp4;
- for audio recordings: mp3, mid, midi, wav, wma;
- for executable file: exe, zip.

Every item can receive tags¹⁵, i.e. non-hierarchically ordered keywords or phrases which classify the contents so that it can be found more easily at

¹⁵ https://omeka.org/codex/Managing_Tags_2.0

Уреди објекат #772: "Модел за оптимизацију гранично..."

Dublin Core Метаподаци врсте објекта Zotero Документи Етикете Везе јединице Мета TEI

Zotero

Dublin Core Метаподаци врсте објекта Zotero Документи Етикете Везе јединице Мета TEI

Метаподаци

Dublin Core Метаподаци врсте објекта Zotero Документи Етикете Везе јединице Мета TEI

Врста објекта **Dublin Core**

Сет елемената метаподатака Даблинско језгро заједничко је за све Омека записе, укључујући објекте, документе и колекције. За више информација видети <http://dublincore.org/documents/dces/>.

Наслов Име дато извору

Модел за оптимизацију граничног садржаја метала у руди у функцији дугорочног планирања површинских копова

Уклони Користи HTML

Model for the Cut Off Grade Optimization in the Function of Long-Term Planning of the Open Pits

Уклони Користи HTML

Аутор Ентитет примарно одговоран за стварање извора

Кржановић Даниел

Користи HTML

Сними измене

Погледај објављену страницу

Обриши

Објављено: ☒ Истакнуто: ☒

Колекција

- Докторске дисертације

Artist [Додај упис](#)

Attorney Agent [Додај упис](#)

Author [Додај упис](#)

Original Format [Додај упис](#)

Physical Dimensions [Додај упис](#)

Figure 2. Describing objects with metadata

Уреди објекат #772: "Модел за оптимизацију гранично..."

Dublin Core Метаподаци врсте објекта Zotero **Документи** Етикете Везе јединице Мета TEI

Документи

Клици и превуци документе у жељени редослед приказа.

DD_Krzanovic Daniel.pdf Уређивање - Обриши

Додај нове документе

Максимална величина документа је 200 MB.

Пронађи документ [Choose File](#) No file chosen

[Додај / рути документ](#)

Add Dropbox Files

Filter files by name:

File Name

☒ DD_Krzanovic Daniel.pdf

☐ Law on Mining and Geological Explorations.pdf

☐ Pravilnik o klasifi i kategor rezervi cvrstih min sir.pdf

☐ Zakon o rudarstvu i geoloskim istrazivanjima.pdf

Сними измене

Погледај објављену страницу

Обриши

Објављено: ☒ Истакнуто: ☒

Колекција

- Докторске дисертације

Figure 3. Panel for the import of documents

a later point. Figure 4 reproduces panels for import of geo-spatial entries, tagging and establishment of relations between items.

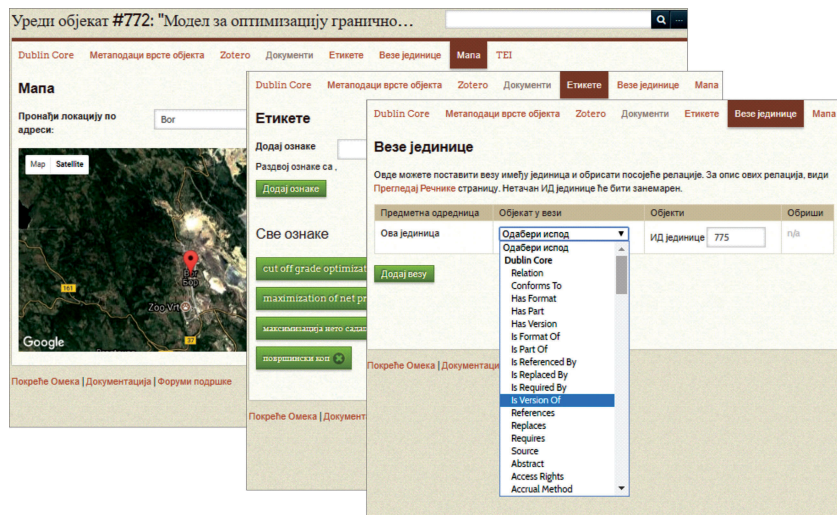


Figure 4. Import of geo-spatial entries, tagging and establishment of relations between items

In the *Map* panel added are the information on the location which is in connection with the item, and it is possible to search items with location in view.

It is possible to establish relation between two or more items. For example, The Law on Amendments and Supplements of The Law on Safety and Health is in the *isPartOf* relation with The Law on Safety and Health.

3.2 Collections

Collections represent groups of items that are organized in a way that makes it easier to search through them. It is meticulously described by meta-data from the expanded set of the Dublin core elements and from the Zotero platform (Figure 5). They can be hierarchically structured, meaning that they can have defined subordinate and superordinate collections. The range of subordinate collections of any one collection spans from none to infinitely

many, while there is always a single superordinate collection. For every collection it is possible to define visibility on the web page.

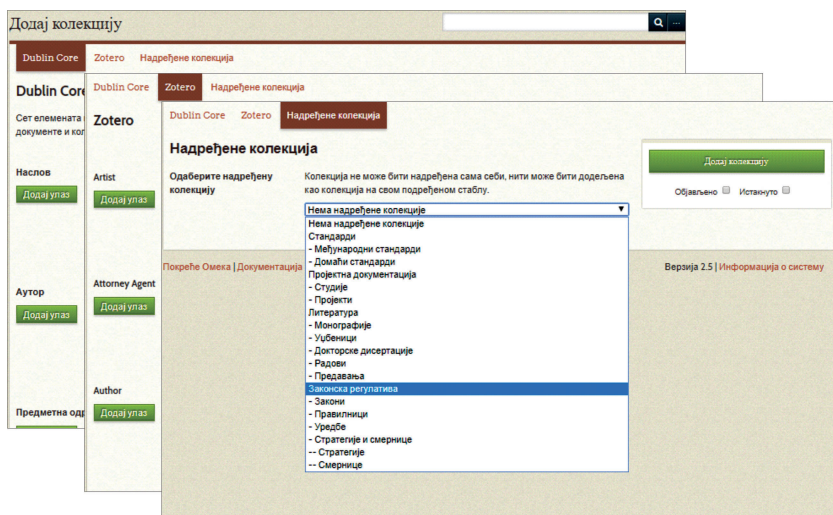


Figure 5. Panels for creating collections

3.3 Web pages

It is necessary to create web pages in order for items and collections to be visible on the web. Apart from the web page title it is necessary to define a relative path to the web page, as a part of the URL-address of that web page (Figure 6). In the part *Text* it is allowed to enter codes from the list of abridged codes, through which the appearance of the webpage itself is adapted. In the example (Figure 6) a code is given for presentation of the collection identified as ID 2 (*Laws*), with the number of items listed on one page limited to 50.

```
[items collection=2
items num=50]
```

After the web page is created, in the part *Navigation*, on the panel for managing the appearance of the web site, the basic data (title and URL)

are entered for every individual web page, while their order is determined through simple dragging of a field to the desired position. In this way, web pages are hierarchically structured and appear on the web site as drop-down menus.

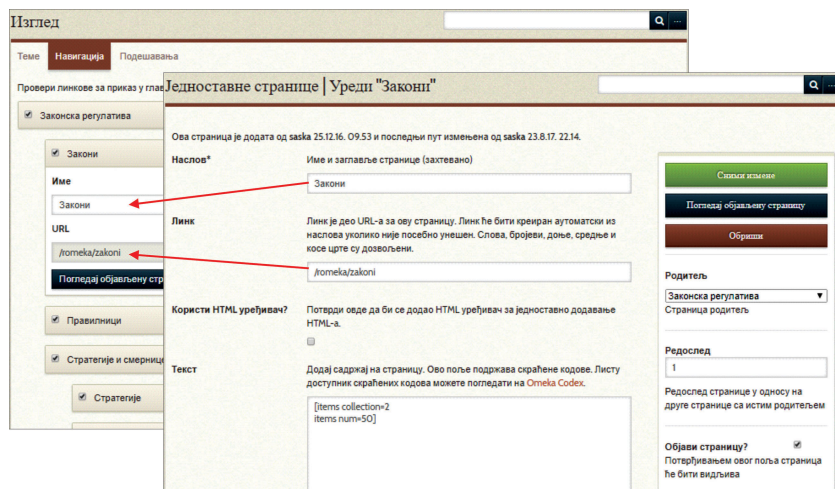


Figure 6. Creation of a web page

4 The digital library ROmeka@RGF

The digital library ROmeka@RGF¹⁶ contains 209 texts primarily in the field of mining, but also in the fields of security, occupational safety, risk assessment, as closely related fields. The selection of documents of the digital library was based on accessible digital resources and for purposes of this digital library no further scanning was arranged for documents previously available in paper format only.

¹⁶ The name ROmeka@RGF was coined by combining the abbreviated name of the Mining department (Rudarski odsek – RO) of the Faculty of Mining and Geology (Rudarsko-geološki fakultet – RGF) and the name of the Omeka web platform.

Items are categorized into 4 main, superordinate collections and 15 subordinate collections. The hierarchical structure of these collections is presented in the Table 1.

Collection	Subordinate Collection	
Legislative	Laws	
	Rulebooks	
	Statutes	
	Strategies and directives	Strategies Directives
Project documentation	Studies	
	Projects	
Literature	Monographies	
	Textbooks	
	Doctoral dissertations	
	Papers	
	Talks	
Standards	International standards	
	Domestic standards	

Table 1. Hierarchical structure of collections

All items are meticulously described by all types of metadata that have been listed in the section 3.1, except for the metadata on use, which we plan to set in motion in the nearest future.

Visibility on web is provided for most of the digital items, but limitations are introduced for all items from collections *Project documentation* and *Standards* and for several items from the collection *Literature*. The reason for this limitation is related to the question of the publication rights, either because of confidentiality or because of author's rights.

All texts that are stored in the digital library will be used for research in terminology. In order to accomplish this, texts have been cleaned and parts in foreign languages have been removed together with tables, images, references and links. In order for these texts to be jointly processed, they have been merged into a single textual file with size of 39 MB, 6200 pages of text of A4 format. The processing of the text yielded 150.365 sentences and 2,719,086 (100,414 different) monomial lexical units. Around 1900 monomial terms, used particularly in the fields of mining, security, occupational safety and

risk assessment are in preparation. After that, polynomial terms will be extracted, using methodology described in the paper (Stanković et al., 2012). We also plan to integrate the search through the corpus of mining texts with the Serbian Language Corpus SrpKor.

5 Search through the textual resources

Search for information in textual resources (comprising of a set of methods and techniques) takes into account both the resources themselves and/or the metadata describing those resources (Baeza-Yates and Ribeiro-Neto, 1999). Systems intended for information search are based on two concepts: *query* and *object*. Queries are formal requests for necessary information entered by the user into the search system. Objects are entities comprising requested information. User queries are matched with objects which are often stored in databases. Examples of the object data are documents and web pages.

Simple search for information in textual resources is implemented as string matching and does not take into account syntactic and semantic features of the requested word. These queries comprise either of one or of multiple words which can be connected through logical operators „and/or“.

When it comes to searching through a digital library, the formulation of more complex queries is enabled through expanded regular expressions. Depending on resources one is searching through, answers can range from documents, to metadata or lists of web pages.

In the course of the search, three basic measures have to be taken into account: *recall*, *precision* and *ranking*. Recall expresses the level of completeness of answers received to a particular query and is represented as the ratio of the total number of relevant documents that were found and the total number of all relevant documents. Precision expresses the level of correctness of answers received to a particular query and is represented as the ratio of the total number of the total number of relevant documents that were found and the total number of answers that were found. Recall indicates the level of comprehensiveness of the system in the course of the search for relevant information.

Problems related to searching through the textual resources can be classified in two categories:

- general problems which are not related to language and

- problems that are specific to a particular language or a group of languages.

The problem encountered in the course of searching through the texts written in Serbian are different code schemes, as well as the existence of two alphabets (the Cyrillic and the Latin alphabet). This is the case with the digital library ROmeka@RGF because documents have been imported in their original form and in both scripts. Search without expansion of the query allows the contents to be searched through on the bases of one alphabet alone, while the expansion of the query automatically involves both alphabets. Searching documents written in Serbian is a complex process because of the very rich morphological system of the Serbian language. Most often, queries are looking for words in their canonical shape (Nominative singular for nouns, infinitive for verbs) (Lazić et al., 2016). However, documents can contain any flection of a flectional word. This problem becomes even more complex if compound words and synonyms are also taken into account (Stanković, 2009).

Elements of a digital library that can be included in the process of searching are: metadata, documents, tags, reports, expositions, web pages. Originally, the search is performed through: key words, bull operators and complete matching. The digital library ROmeka@RGF is upgraded through implementation of expanded queries. Web services (Stanković et al., 2012) and morphological electronic dictionaries for Serbian (Krstev et al., 2008; Stanković et al., 2016) have been used:

- for a morphological expansion of a query::
http://hlt.rgf.bg.ac.rs/vebran/api/delafs/ključna_reč
- for a semantic and morphological expansion of a query:
http://hlt.rgf.bg.ac.rs/vebran/api/sinonimi/ključna_reč

Searching with and without morphological and semantic-morphological expansions of queries will be presented on the example of a search for a lexeme *хомогенизацѝја*.

An unexpanded query is taking only the Nominative singular into account:

хомогенизацѝја

A query with a morphological expansion allows the inclusion of all flectional forms of the requested lexeme, on both alphabets (Latin and Cyrillic). Therefore, in this case, the following flections are searched through:

homogenizacija, homogenizacijama, homogenizacije, homogenizaciji, homogenizaciju, homogenizacijom, homogenizaciju, хомогенизација, хомогенизацијама, хомогенизације, хомогенизацији, хомогенизацију, хомогенизацијом, хомогенизацију.

A query containing both semantic and morphological expansions allows the inclusion of synonyms and lexical metonymies, as well as of their flexional forms:

homogenizacija, homogenizacijama, homogenizacije, homogenizaciji, homogenizaciju, homogenizacijom, homogenizaciju, уједначавање квалитета угља, homogenizacija квалитета угља, homogenizacijama квалитета угља, homogenizacije квалитета угља, homogenizaciji квалитета угља, homogenizaciju квалитета угља, homogenizacija угља, homogenizacijama угља, homogenizacije угља, homogenizaciji угља, homogenizaciju угља, homogenizacijom угља, homogenizaciju угља, управљања квалитетом угља, управљање квалитетом угља, управљањем квалитетом угља, управљањима квалитетом угља, управљању квалитетом угља, управљања квалитетом, управљање квалитетом, управљањем квалитетом, управљањима квалитетом, управљању квалитетом, хомогенизација, хомогенизацијама, хомогенизације, хомогенизацији, хомогенизацију, хомогенизацијом, хомогенизацију, уједначавање квалитета угља, хомогенизација квалитета угља, хомогенизацијама квалитета угља, хомогенизације квалитета угља, хомогенизацији квалитета угља, хомогенизацију квалитета угља, хомогенизацијом квалитета угља, хомогенизацију квалитета угља, хомогенизација угља, хомогенизацијама угља, хомогенизације угља, хомогенизацији угља, хомогенизацију угља, хомогенизацијом угља, хомогенизацију угља, управљања квалитетом угља, управљање квалитетом угља, управљањем квалитетом угља, управљањима квалитетом угља, управљању квалитетом угља, управљања квалитетом, управљање квалитетом, управљањем квалитетом, управљањима квалитетом, управљању квалитетом.

Results of the search for monomial and polynomial terms (*хомогенизација, управљање квалитетом угља, површински коп, роторни базен, експлоатација и рударство*) in the digital library Romeka@RGF, figure in the Table 2.

The diagram (Figure 7) illustrates the results of the search for monomial and polynomial terms in the digital library Romeka@RGF based on the data

Type of query	homogenizacija	upravljanje kvalitetom uglja	površinski kop	rotorni bagger	eksploatacija	rudarstvo
Without expansion for the Cyrillic script	6	0	21	0	13	15
Without expansion for the Latin script	9	14	49	33	50	23
Morphological expansion	33	22	106	42	120	106
Semantic and morphological expansion	47	47	118	43	123	120

Table 2. Results of the search for monomial and polynomial terms

given in the Table 2. It is worthwhile observing that queries with morphological and semantical expansions yield more results than unexpanded queries.

6 TEI (Text Encoding Initiative)

In order to extract information from digital items which are a part of the digital library ROMeka@RGF, we have decided to store items which are in the textual format in the XML format as well in accordance with the TEI¹⁷ guidelines. We have made this decision because TEI is a *de facto* standard for annotation of arbitrary document types, including legal texts and project documentation texts. Annotation in accordance with TEI guidelines should enable the connection between parts of a project documentation text which refer to law articles and legal regulations.

In the course of annotation in accordance with the TEI P5 guidelines (TEI-Consortium, 2017), tags <div1>, <div2>, <div3> and <div4> were used in accordance with the hierarchy between a law, a chapter, a section and an article (Васильев, 2015). The tag <p> was used as equivalent to point, subpoint and indent of a law.

Tags with different attribute values were used for tagging of titles and subtitles. In accordance with the connection prerequisites and mindful of

¹⁷ <http://www.tei-c.org/>

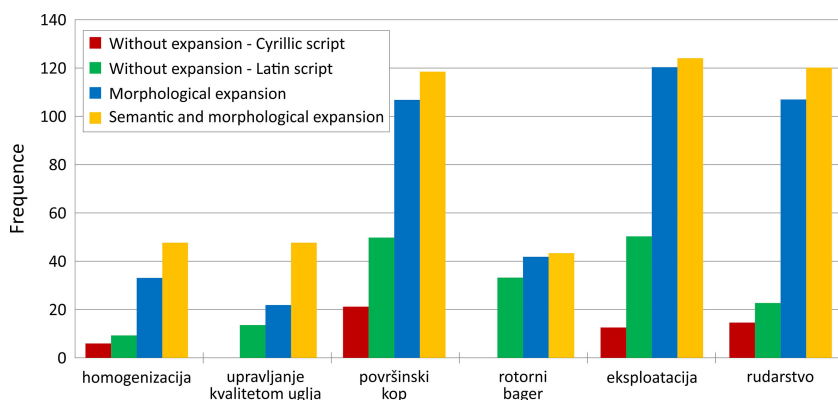


Figure 7. Results of the search for monomial and polynomial terms in the digital library ROmeke@RGF

the connections between the law articles, we used the tag `<head>` to mark the title, chapter, section and article of a law. Within this tag, attributes *type* and *n* were used. The values of the attribute *type* are crucial for the distinction between different parts of a law. There are four possible values of this attribute: *main*, *chapter*, *section* or *article*. The values of the attribute *n* are the ordinal numbers of the current elements (i.e. of a chapter, section or article).

Figure 8 represents a part of the document *Law on Mining and Geological Explorations* tagged in accordance with the TEI guidelines in the way discussed previously.

Mindful of the mining project presentations, we made an effort to conserve all elements – most notably the tables – as important sources of information, while we were using TEI. Figure 9 illustrates a tabular presentation within the TEI version of the project documentation.

7 Conclusion

In this paper both advantages and disadvantages were shown of Omeka as a platform for development of this type of library. Additionally, the extent to which the application of morphological dictionaries affects the quality of a search in the course of a morphological and semantical expansion of a query

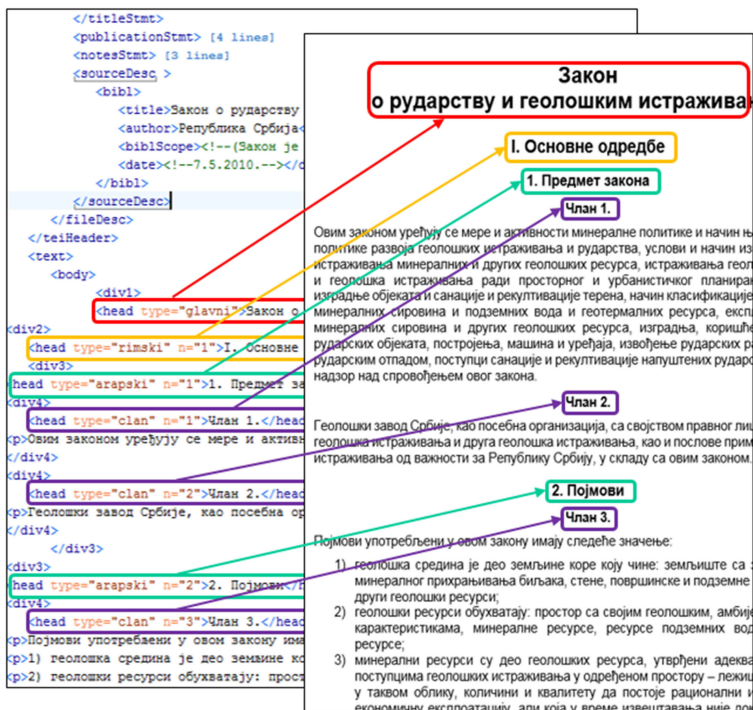


Figure 8. A part of the document tagged in accordance with the TEI PS guidelines

Tabela 1.5.1.	
Parametar	Vrednost
Vlaga, %	39,22
Pepeo, %	17,70
S ukupni, %	1,18
S sagorljiv, %	0,56
S u pepelu, %	0,60
Koks, %	34,98
C-f ix, %	18,26
Isparljivo, %	26,30
Sagorljivo, %	43,14
Gornja toplota sagorevanja, kD/kg	11.490
Donja toplota sagorevanja, kJ/kg	10.020

Figure 9. TEI P5 illustration of the tabular presentation

has also been shown. The conclusion is that combining Omeka and morphological dictionaries is likely to assist better organisation and searchability of items stored in the digital library ROMeka@RGF.

Since it is important that both the collections and the search tools are regularly upgraded, we plan to follow the developments of further plugins and technologies which could assist the connection of the digital library ROMekaRGF with other sources of similar information. Furthermore, we plan to supplement the digital library with various new digital items. We also plan to extract information from digital items themselves.

Acknowledgements

The research presented in this paper has been financially supported by the Ministry of education, science and technological development of the Republic of Serbia through the projects TP33039 ("Improvement of Lignite Opencast Mining Technology in Order to Increase Energy Efficiency and Occupational Safety").

References

- Arms, William Y. *Digital libraries*. Cambridge, Massachusetts, USA: M.I.T. Press, 2000. <http://www.cs.cornell.edu/wya/diglib/ms1999/>
- Baeza-Yates, Ricardo and Berthier Ribeiro-Neto. *Modern information retrieval*, Vol. 463, New York, USA: ACM Press, 1999. <http://web.cs.ucla.edu/~miodrag/cs259-security/baeza-yates99modern.pdf>
- Hodge, Gail. *Metadata made simpler*. Niso Press, 2001.
- Krstev, Cvetana, Ranka Stanković, Dusko Vitas and Ivan Obradović. "The Usage of Various Lexical Resources and Tools to Improve the Performance of Web Search Engines". In *LREC*, Paris, France: European Language Resources Association (ELRA), 219–224, 2008. http://lrec-conf.org/proceedings/lrec2008/pdf/67_paper.pdf
- Kucsma, Jason, Kevin Reiss and Angela Sidman. "Using Omeka to build digital collections: The METRO case study". *D-Lib magazine* Vol. 16, no. 3/4 (2010). <http://webdoc.sub.gwdg.de/edoc/aw/d-lib/dlib/march10/kucsma/03kucsma.html>
- Lazić, Biljana, Danica Seničić, Aleksandra Tomašević and Bojan Zlatić. "Terminological and Lexical Resources Used to Provide Open Multilingual Educational Resources". Belgrade, Serbia, 2016. http://www.baektel.eu/documents/conferences/eLearning_2016_BL_DS_AT_BZ.pdf

- Stanković, Ranka, Cvetana Krstev, Ivan Obradović, Aleksandra Trtovac and Miloš Utvić. "A tool for enhanced search of multilingual digital libraries of e-journals". In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC)*, Istanbul, Turkey, 1710–1717, 2012. http://www.lrec-conf.org/proceedings/lrec2012/pdf/375_Paper.pdf
- Stanković, Ranka, Cvetana Krstev, Biljana Lazić Ivan Obradović and Aleksandra Trtovac. "Rule-based Automatic Multi-Word Term Extraction and Lemmatization". In *Proceedings of the 10th International Conference on Language Resources and Evaluation, LREC*, 507–514, 2016. http://www.lrec-conf.org/proceedings/lrec2016/pdf/1033_Paper.pdf
- Stanković, Ranka M. "Modeli ekspanzije upita nad tekstuelnim resursima". Doktorska disertacija. Univerzitet u Beogradu, Matematički fakultet, 2009.
- TEI-Consortium. "TEI P5: Guidelines for Electronic Text Encoding and Interchange. Version 3.2.0.", 2017. Accessed August 23, 2017, <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/>
- Васиљевић, Небојша. "Аутоматска обрада правних текстова на српском језику". Докторска дисертација. Универзитет у Београду, Филолошки факултет, 2015. <https://fedorabg.bg.ac.rs/fedora/get/o:10687/bdef:Content/get>
- Тртовац, Александра С. "Дескриптори метаподатака и дескриптори садржаја у проналажењу информација у дигиталним библиотекама". Докторска дисертација. Универзитет у Београду, Филолошки факултет, 2016. <https://fedorabg.bg.ac.rs/fedora/get/o:12605/bdef:Content/get>

eLibrary of the Goethe-Institut

UDC 004.738.5:027.54

DOI 10.18485/infotheca.2017.17.2.3

Biljana Pajić

biljana.pajic@goethe.de

Goethe-Institut Belgrade

ABSTRACT: The first chapter of this paper presents theoretical terms in the field of eLibraries, which are later further explained using the example of the eLibrary of the Goethe-Institut, called Onleihe. After a short historical overview of eLibraries in general, the final chapter presents the results of the survey conducted among active users of the Onleihe. The purpose of this survey was to determine users' habits as well as their impressions considering the offer of the Onleihe. We also tried to determine the causalities between their gender, age, education level, location, knowledge of the German language and previous experiences in traditional and eLibraries, if any, on the one hand and the aforementioned habits and impressions on the other.

KEYWORDS: eLibrary, digital library, digital reading.

Translated by:

Svetlana Milivojević-Petrović

PAPER SUBMITTED: 4 September 2017

PAPER ACCEPTED: 4 December 2017

1 E-libraries

1.1 Definition of terms

In order to fully understand the subject matter relating to electronic libraries, it is necessary to define first the terms such as electronic books, electronic reader, electronic library and digital rights management (*DRM: Digital Rights Management*).

According to the electronic version of the Oxford dictionary, an electronic book, abbreviated as eBook, is an electronic version of a printed book that can be read on a computer or a handheld device (mobile phone), designed

specifically for that purpose.¹ In the older editions of the Duden and Brockhaus encyclopaedias, we find the definitions of an eBook (*e-Book*, *elektronisches Buch*) that have in the meantime been ascribed to the term electronic reader (*e-Reader*). Namely, an eBook is defined there as a reading device the size of a book, which displays a book on an LCD screen. However, in the Brockhaus encyclopedia, it is implied that the database of a book which is read through that kind of device is often referred to as an electronic book.

Because of the obvious problem such terminology could lead to, the term “electronic reader” has been introduced. By that definition, an electronic reader would be a mobile device the size of a book which is used for displaying electronic books (Duden, 2007; Brockhaus, 2003).

“Electronic” or “digital library” is a term which created a lot of terminological confusion towards the end of the 20th century, alongside the related terms such as “virtual“, “hybrid” or “a library without walls“. From a number of various discussions related to this topic, Cleveland highlighted several characteristics of digital libraries, on the assumption that they ultimately remain – libraries. To him, digital libraries represent the face of traditional libraries and include both digital and printed material, as well as all the processes and services that constitute the nervous system of a library (Cleveland, 1998). Nevertheless, Borgman defines digital libraries as a set of electronic resources and technical possibilities for their creation, search and use. According to Borgman, digital libraries include data, metadata which describe various aspects of data, and metadata which consist of links or connections to other data or metadata, regardless of whether they are internal or external links (Borgman, 1999). Through the analysis of 64 definitions collected within a student project, Schwarz singled out several functions of digital libraries, which could be ascribed to them in an ideal situation. Thus she concludes that a digital library is a separate set of electronic resources and activities, which could, but do not have to be organized within the operational system of a traditional library (Schwartz, 2000)

Besides these three terms, it is also very important to define the term digital rights management. There is still no universal definition of *DRM* (Fränkl and Karpf, 2004), but we are going to take into consideration Ianella’s definition, which is one of the most cited ones. According to this definition, *DRM*, namely, comprises description, segmenting, analysis, evaluation, trade in and monitoring of the rights of a certain enterprise and

¹ Accessed 01.09.2017, <https://en.oxforddictionaries.com/definition/us/e-book>

pertains – in the context of publishing – to both physical and digital published books and other media (Ianella, 2003). *DRM* also pertains to data protection against copying, depending on the previously defined copyrights. The website *Ebook Architects* provides an illustrative description of *DRM* as being a lock attached to a digital database connecting it to the account of a person who has purchased it, i.e. rented it. In addition, the experts from the same website criticize *DRM* as an outdated manner of book protection, but they do not deny that this kind of protection is necessary. However, they hold the view that a new way of copyright protection of digital databases is necessary.

1.2 Format review

Despite efforts along these lines, there are still no universal standards for eBook formats, nor do they exist as regards the devices and software which are used to display them. Nevertheless, there are several formats which are most frequently used in electronic publishing. We are going to use the review provided by the website *Ebook Architects*.²

Kindle Format 8 (KF8) – this is the format used by the company Amazon, adjusted to the electronic readers Kindle, as well as Kindle-applications. This format allows book locking by means of *DRM*, linking it exclusively to the device the book has been loaded onto (Figure 1, left).



Figure 1.

ePub2 / ePub3 – these formats have been developed as a general standard for eBooks. They are based on different technologies and standards, and are unique in terms of how they are combined. *DRM* is most effectively integrated into these formats through *Adobe Content Server* program. *NOOK Kids* is a very similar format to these, and is used for displaying children’s books (Figure 1, middle and right).

² Accessed 01.09.2017, <http://ebookarchitects.com/learn-about-ebooks/formats/>

iBooks Author – this format has been created by the *Apple* company, and has been mainly designed for non-fictional electronic books. It is not possible to build a standard *DRM* into the books using format, but the Apple company has created its own technology called *FairPlay*, which serves the same purpose (Figure 2, left).



Figure 2.

PDF – this format is certainly the most widespread, despite gradually becoming obsolete in the modern eBook world, mostly because it cannot be adjusted to the small screens of electronic readers, so, although readers in general support this format, reading is not so “comfortable“, as is the case with the aforementioned formats. Concerning digital rights, these databases can be encoded or locked (Figure 2, right).

1.3 Historical overview

Candela, Casteli and Pagano refer to the invention of Vannevar Bush from the year 1945 as a precursor of electronic libraries – a device which a person can use to store all their books, files and correspondence, and which can function flexibly and at great speed. One of the visionaries they mention is Lickleider, who in 1965 recognized the potential of computer technologies to create an automatic library which can be used by a number of persons simultaneously. However, serious research in this field started in the 1990’s (Candela et al., 2011).

According to Besser, the first major acknowledgment of the importance of electronic libraries occurred in 1994 by the announcement that 24.4 million American dollars from the US federal budget would be spent on research in electronic libraries, involving six universities (Besser, 2004). Although he considers them very exciting, Besser nonetheless emphasizes that the results of these researches were a far cry from libraries, consequently deeming this phase “experimental“. He subsequently introduces us to the so-called “development” phase, which ensued four years later. During this phase the American Government invested money in a project which served the purpose

of adding some of the services of a traditional library to this experimental product, while at that time, he states, the administrators of traditional libraries started to develop serious digital components.

Various initiatives appeared, which spawned research projects in this field, both in the US and Europe (Besser, 2004). When librarians took part in these projects to a larger extent, the attempts at creating electronic libraries advanced from informatics experiments to more operational services.

Nowadays all initiatives and projects can be divided into three groups: those which are guided by the principle of data sharing, those committed to the definition and development of generic systems for the purpose of simplifying the construction and operation of electronic libraries, as well as those leading to new research environments (Besser, 2004).

2 About the Goethe-Institut

The Goethe-Institut is a cultural institution of the Federal Republic of Germany which operates worldwide. It supports and promotes study of the German language abroad and fosters international cultural cooperation. It also conveys a comprehensive image of Germany by providing information about its cultural, social and political life. With the network of Goethe-Instituts, Goethe Centres, cultural societies, reading rooms, exam and language learning centres, it assumes the central task of the educational and cultural policy of the Federal Republic of Germany abroad. It also establishes partnerships with public and private cultural institutions, federal provinces, communes and the economy. The Goethe-Institut in Belgrade was founded in 1970. It is responsible for the cultural exchange between Germany on the one hand and Serbia and Montenegro on the other.

The Goethe-Institut in Belgrade is a part of the regional network of the Goethe-Instituts in South-Eastern Europe, and in that respect it provides strong support to the implementation of regional projects, topics and program goals. The work of the Goethe-Institut in South-Eastern Europe is coordinated by the institute in Athens, and apart from the institute in Belgrade, the other institutes in this region are in Ankara, Bucharest, Istanbul, Ismir, Nicosia, Sarajevo, Sofia, Thessaloniki, Tirana and Zagreb.

The activities of the Goethe-Institut in Belgrade include the following spheres of work:

- cultural programs;
- providing information and library services;

- language courses and corresponding exams;
- cooperation in language teaching.³

2.1 About the Goethe-Institut eLibrary

eLibrary is the Goethe-Institut's digital library. This service enables customers to borrow digital media on a time-limited basis by simply downloading them from eLibrary.⁴

The eLibrary services on offer by the Goethe-Institut have been provided by the company *divibib GmbH*⁵ from Wiesbaden in Germany, a daughter company of *ekz.bibliotheksservice GmbH*. Since 2007, the company *divibib* has offered one of the leading platforms of electronic libraries called *Onleihe*⁶ and it currently boasts more than 3,000 libraries among its clients. This platform is used in Germany by all major public libraries, and also by libraries in Denmark, Belgium, France, Italy, Liechtenstein, Austria and Switzerland. In addition to the standard services offer through the web pages of libraries that use the services of the company *divibib*, there is a possibility of using applications for *Android* and *iOS* platforms free of charge.

In 2011, the network of Goethe-Instituts included eLibrary among the services it offers, and even in that initial phase some of the libraries from South-Eastern Europe joined in. The eLibrary of the Goethe-Institut currently contains 19,225 units,⁷ which comprise popular fiction, books for children and youth, manuals and handbooks from various specialist fields, as well as music, audio books and movies. The current services on offer include electronic books and magazines in *pdf*, *ePub* and *.acsm* formats, and audio material (music or audio-books) in *.wma* format, whereas the movies on offer can only be watched online. The library holdings are available to all the libraries of the Goethe-Institut which are included in the network of electronic libraries, but terms of service vary from one institute to another. Unlike the majority of other institutes, the Belgrade Institute enables its users to use this service free of charge, just as is the case with the services offered by the traditional library.

³ Accessed 01.09.2017, <https://www.goethe.de/ins/cs/sr/ueb/auf.html>

⁴ Accessed 01.09.2017, <https://www.goethe.de/ins/cs/sr/kul/bib/onl.html>

⁵ Accessed 01.09.2017, <http://www.divibib.com/>

⁶ Accessed 01.09.2017, <http://www.onleihe.net/>

⁷ Accessed 01.09.2017, <http://www2.onleihe.de/goethe-institut/frontend/welcome,51-0-0-100-0-0-1-0-0-0-0.html>

The users register filling out a form on the web page of the Goethe-Institut⁸ in Belgrade, providing their basic data such as their name, surname, date of birth, address, phone number and their library membership card number of the Goethe-Institut, if they have one (Figure 3, left).

Figure 3.

This registration reaches the library by means of an email, and the librarian subsequently generates a unique activation code from his or her account on the portal *Mein goethe.de* and sends it by email to the user with the accompanying instructions on how to use it in PDF format. Then the user, if he or she has not done it already, opens an account on the portal *Mein goethe.de*, where in the section “eLibrary” he or she enters the received activation code. That is how an account is activated for the duration of one year. Upon the expiration of that period, the user may register again following the same procedure (Figure 3, right and Figure 4).

The eLibrary holdings can be searched by a simple searching method, or browsed through the categories they are divided into. When the user finds the desired medium, it is necessary to “place it into the library basket” first, where he or she can select a suitable format and download the database. It is not necessary to return the borrowed databases. After the expiry of the loan period, the database simply cannot be opened any more.

⁸ Accessed 01.09.2017, <https://www.goethe.de/ins/cs/sr/kul/bib/onl/anm.html>



Figure 4. eLibrary visual identity (design: Slavimir Stojanović Futro)

The Goethe-Institut in Belgrade has been in charge of the eLibrary services on offer since January 2014. Since then – for more than three and a half years – 2,600 activation codes have been generated, which makes up for more than 10% of the overall number in the region of South-Eastern Europe. Out of that number, only 1,315 users have activated their account, whereas 1,285 of them – less than a half – have not used the registration code they received. There are indications this percent is due to a somewhat complicated registration process, which requires form filling, waiting for a registration code, which is not automatically generated, but is granted manually, registration at the portal *Mein goethe.de* and finally account activation. According to subjective assessments, there are too many steps to take in order to register, so less savvy computer users are frequently demotivated before they carry out the account activation. In some cases, users do not provide their correct email address, so the librarian is unable to forward their registration code to them. However, in July 2017, the traditional library system in the physical library was replaced by the modern software Koha, which additionally enables users to keep their personal search requests, extend loan periods and reserve media from the library. These personalized accounts are available to the library users as of November 2017. When this practice becomes more common, a gradual increase of eLibrary activated accounts is expected, taking into consideration that the physical library account is activated via the portal *Mein goethe.de*, the same portal which is used to log into the eLibrary.

The eLibrary is currently⁹ being used by 338 active users from Serbia, which makes up for 16.1% of the overall number of active users in South-Eastern Europe. In 2014, there were 2.22 users registering per day, while in 2015 the average percentage was somewhat lower: 1.81. In the year 2016 there were 2.12 of users on average who registered with the eLibrary per day.

⁹ Data from August 2017.

3 Survey analysis

"The Goethe Institut eLibrary in Belgrade" survey was conducted from September 2015 to August 2017 on a sample of 327 respondents, and new data are constantly being collected. The aim of the survey was to establish the respondents' user habits, their impressions in regard to the offer and user services of the eLibrary, as well as the causalities between their gender, age, education, place of residence, knowledge of the German language and previous experience in electronic and traditional libraries on the one hand, and their habits and impressions on the other.

3.1 Statistical descriptors

The number of respondents according to gender is practically even. Namely, 53.2% of the respondents are women and 46.8% men (Chart 1, left).

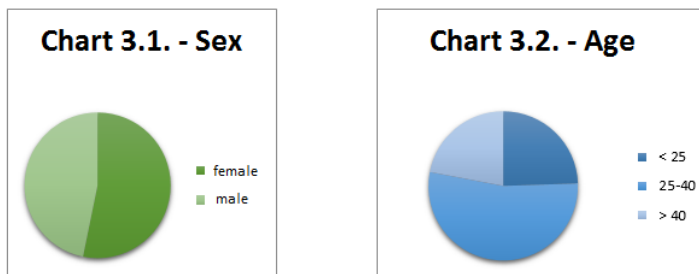


Chart 1.

As regards the age of the respondents, according to the sample we can conclude that the majority of active eLibrary users are aged between 25 and 40. 24.5% of the respondents are under 25, 53.5% are between 25 and 40 years, and 22.0% above 40 (Chart 1, right).

According to the place of residence, we divided the results into two groups: the respondents living in Belgrade and those who live outside Belgrade. The results show that these two groups are almost the same. Namely, 57% of the respondents live in Belgrade, while 42.8% reside in other cities in Serbia (Chart 2, left).

34.3% of the respondents graduated from high school or a place of higher education, 28.1% completed their basic university studies while the rest of them, a sizeable 37.6%, obtained a Master's or PhD degree (Chart 2, right).

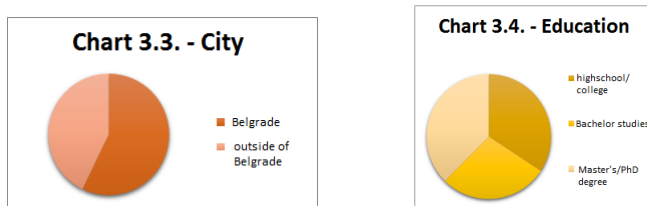


Chart 2.

According to these data, we conclude that the majority of the eLibrary users of the Goethe-Institut are highly educated, which can be explained by their high degree of computer literacy which academic education implies.

When they were asked to assess their knowledge of the German language on a scale from 1 to 5, the lowest grade being 1 and the highest 5, the respondents' answers were as follows: a mere 12.5% of the respondents assessed their knowledge with the lowest grade, 17.4% would give themselves grade 2, 33.3% assessed their knowledge with an average grade 3, 22.0% would give themselves grade 4 while 14.7% of the respondents think they possess an excellent command of the German language. Therefore we conclude that the majority of users of the Goethe-Institut eLibrary have a solid to excellent knowledge of the German language. Given that the whole content of the eLibrary is in the German language, this result was easy to predict. However, taking into consideration that the eLibrary interface can be used in English as well, or some other world language, and that the library holdings include material for the study of the German language, it is no surprise that there are, among active users, those who do not assess their knowledge of German with the highest grades (Chart 3, left).

The greatest number of the respondents, 27.2% of them, use the eLibrary several times per month, while 42.5% do so once or several times per month (Chart 3, right).

42.5% of the Goethe-Institut eLibrary users simultaneously use the Institute's "paper" library, whereas the rest of the respondents, 57.25%, if we compare them with the data concerning their place of residence, can be

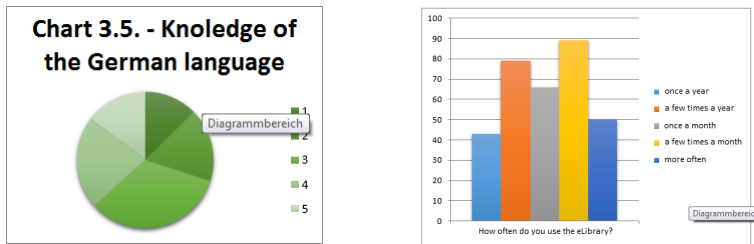


Chart 3.

explained by the fact that the Goethe-Institut does not have its branches outside Belgrade. That was later confirmed by using an independent samples t-test, where it was found there was a considerable difference between the number of users of the traditional library of the Goethe-Institut, who reside in Belgrade, and those living in some other city – to show that those from Belgrade prevail. The majority of the respondents, 60.9% of them precisely, simultaneously hold membership cards of some other libraries, while only 28.1% use an additional eLibrary. The latter is explained by the fact that ePublishing has not gained momentum in Serbia yet and, as a consequence, there is no substantial offer of electronic library holdings (Chart 4, left).

There follows a set of questions the respondents assessed on a scale from 1 to 5, depending on the extent to which they agree with a statement. In the first group of questions, there are those dealing with application and registration in order to use the Goethe-Institut eLibrary. According to the results, 70.9% of the respondents find information on the eLibrary easily or quite easily, while not a single respondent finds it extremely hard to obtain information on the eLibrary. As many as 54.7% are of the opinion that sending a request for the use of the e-Library is quite simple, and 60.9% find the instructions on how to register completely clear. 80.7% consider the registration process as quite or absolutely simple. 52.0% assessed the e-Library technical support as fully functional. The majority of the respondents, therefore, faced no obstacles with application or registration. However, taking into account that the survey was conducted among the active users – those who had successfully gone through the process – this result does not come as such a surprise (Chart 4, right).

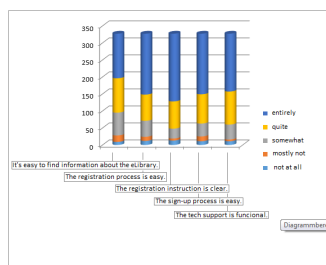
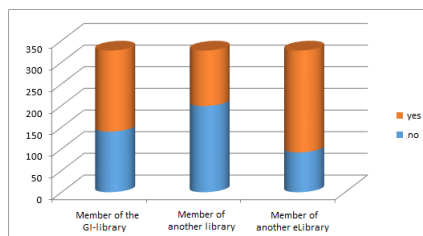


Chart 4.

The other set of questions deals with the user experience itself as regards the Goethe-Institut eLibrary. In this case as well, the majority of the respondents are satisfied with the eLibrary. As many as 74.0% find the required reading material easily or quite easily, while 70,0% of the respondents find the library holdings fully or quite well stocked. 23.9% of them assess the library holdings with an average grade, though. Almost four-fifths of the users (79.5%) are quite or completely satisfied with the efficiency of the media reservation process, while 73.1% of them consider the format selection absolutely or quite great. The vast majority of the respondents (78.9%) assess the media descriptions of the eLibrary as sufficiently informative. (Chart 5)

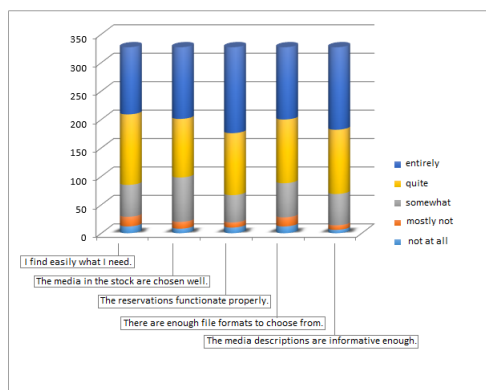


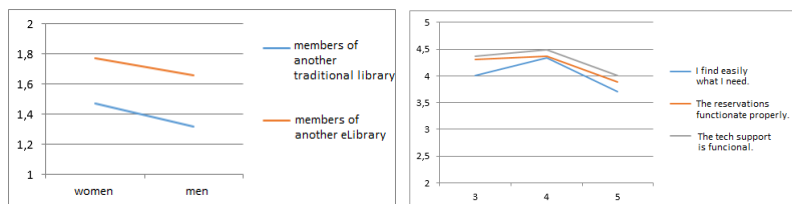
Chart 5.

The respondents were given space to leave additional comments in this survey, where they generally praised the electronic offer and expressed gratitude for being able to use it free of charge. Constructive criticism was mainly related to the eLibrary holdings choice, the majority of the users expressing their wishes to see movies in the German language in the electronic offer, which has been accomplished in the meantime. A demand for academic material was expressed, in the form of specialist literature, outstanding doctoral theses, etc. The eLibrary holdings of specialist literature exist and are renewed at regular intervals by the replenishment of the eLibrary holdings, but content such as academic papers is not compatible with the Goethe-Institut profile and purpose. There is, also, a great demand for the Goethe-Institut exam preparation material, but the publishing houses dealing with this sort of material have not enabled access to these editions in an electronic format yet. In addition, there is a demand for regular replenishment of popular fiction holdings, audio books and specialist literature for German scholars and teachers of the German language, which is the regular eLibrary practice.

When it comes to using eLibrary mobile applications, more than two-thirds of the respondents (68.2%) do not use any of the available applications. 23.5% use the Android application, 6.1% the one for the iOS- platform, and a mere 2.1% use both applications. Among those who use them, 77.3% are quite or very satisfied with their efficiency.

3.2 Comparison groups

The results related to the use of other traditional and electronic libraries in men and women were compared by an independent samples t-test. An independent samples t-test is used to compare the average values of some continuous variable in two different groups of subjects (Pallant, 2009). A considerable difference was established in both cases, in favour of men (Chart 3.2, left).



The survey on the use of some of the applications for the eLibrary in relation to the respondents' gender produced similar results in some earlier researches; this time, however, no considerable difference between genders was discovered. With regard to the application use, a considerable difference was determined in relation to a place of residence – Belgrade citizens use them much more than their counterparts in other cities.

An independent samples t-test established there was a considerable difference in education among the respondents who stated they were members of some other electronic library and those who stated the opposite. Namely, the users of other electronic libraries have a higher degree of education than the ones who do not use those libraries.

By applying one factor analysis of variance, the impact of age, knowledge of the German language, the frequency of the eLibrary use, as well as use of some of the applications for handling the registration process effectively and the user experience were researched respectively. The analysis of variance is used to compare the variability of the results among different groups with a variable within each group (Pallant, 2009). Despite the fact that considerable differences among age groups were found before, that was not the case this time. The knowledge of the language, nevertheless, has a considerable impact on effective use of the eLibrary, although not to the extent it may have been anticipated at first. Namely, the respondents who assessed their knowledge of the German language with the highest grade (5) were less satisfied with the offer than those who assessed their knowledge with grades 3 and 4. We obtained the data by comparing cumulative answer results on a scale of 10 questions, but also for separate answers relating to the question of how easy it is to find the material and the effectiveness of the reservation system, with a considerable difference occurring among those with a very good and excellent knowledge of the German language concerning the technical support satisfaction level (Chart 3.2, right).

We discovered a considerable difference by comparing the answers to the question pertaining to the effectiveness of the reservation system with those to the question of whether the respondents used mobile applications, and if so, which applications. We obtained the results showing *iOS* platform application users were more satisfied with the effectiveness of the reservation system compared to other users.

4 Conclusion

In this paper, we primarily presented the terms related to electronic libraries: electronic books, electronic readers and digital management rights. We made a reference to the most widespread formats of electronic books used today, as well as eLibrary history. Subsequently, we said something more about the Goethe-Institut cultural institution and its Belgrade office, further providing a more in-depth description of the Goethe-Institut offer of services, complementing it with pictures and general statistical data.

We summed up the paper by the survey results analysis, conducted among the active Goethe-Institut eLibrary users. The survey results showed considerable differences in several instances. Although our sample in relation to gender was almost the same, men use other traditional and electronic libraries more. The eLibrary applications are more frequently used by Belgrade citizens compared to other citizens in Serbia. Users with a higher degree of education use other electronic libraries more frequently, which comes as no surprise, bearing in mind that electronic libraries are a very important source for scientific research in the world's academic environment. We did not manage to provide an adequate explanation with regard to the findings that users who assessed their knowledge of the German language with the highest grade were less satisfied with their user experience than those assessing their knowledge with somewhat lower grades.

We conclude that the users are, generally speaking, quite satisfied with the Goethe-Institut eLibrary, although there is room for improvement when it comes to making the registration and activation process easier and replenishing the holdings with video material and didacticized video-material.

For further research, we recommend conducting a survey on a larger sample, so that the results could be as reliable as possible. The Goethe-Institut in Belgrade continues to compile answers from its active users and analyze them annually. Conducting a survey in other Goethe-Institut libraries in South-Eastern Europe could contribute to a broader perspective and would give us insight into the differences caused by different environments.

References

- Besser, Howard. *The Past, Present, and Future of Digital Libraries*, 2004. Accessed August 24, 2017, <http://digitalhumanities.org:3030/companion/view?docId=blackwell/9781405103213/9781405103213>.

- [xml&chunk.id=ss1-5-6&toc.depth=1&toc.id=ss1-5-6&brand=9781405103213_brand](#)
- Borgman, Christine L. “What are digital libraries? Competing visions”. *Information Processing and Management* Vol. 35 (1999): 232–235. Accessed August 29, 2017, <https://pdfs.semanticscholar.org/9f94/1e73f717613fa558acd2c5858691aa6ff76a.pdf>
- Brockhaus. *Computer und Informationstechnologie. Hardware, Software, Multimedia, Internet, Telekommunikation*. Leipzig – Mannheim: F. A. Brockhaus GmbH, 2003.
- Candela, Leonardo and Donatella Castelli and Pasquale Pagano. *History, Evolution and Impact of Digital Libraries*, 2011. Accessed August 25, 2017, http://www.nmis.isti.cnr.it/candela/Leonardo_Candela_Website/my_Papers/Entries/2011/1/1_History,_Evolution,_and_Impact_of_Digital_Libraries_files/History,_Evolution,_and_Impact_of_Digital_Libraries.pdf
- Cleveland, Gary. “Digital Libraries: Definitions, Issues and Challenges”. *UDT Occasional Paper* Vol. 8 (1998): 1–3. Accessed August 29, 2017, <https://pdfs.semanticscholar.org/e0b2/17a53e8f6a1a834790e2a145a87ea2c56311.pdf>
- Duden. *Das Fremdwörterbuch*. Vol. 5. Mannheim: Bibliographisches Institut & F. A. Brockhaus AG, 2007.
- Fränkl, Gerald and Philipp Karpf. *Digital Rights Management Systeme Einführung, Technologien, Recht, Ökonomie und Marktanalyse*. München: PG Verlag, 2004.
- Ianella, Renato. *Open Digital Rights Language (ODRL)*, 2003. Accessed August 25, 2017, <https://www.w3.org/TR/odrl/>
- Pallant, Julie. *SPSS – Survival Manual. A Step by Step Guide to Data Analysis Using SPSS for Windows*, 2009. Translation of the third edition: Miljenko Šućur
- Schwartz, Candy. “Digital Libraries: An Overview”. *The Journal of Academic Librarianship* Vol. 26, no. 6 (2000): 385–386. Accessed August 29, 2017, <https://home.kku.ac.th/hslib/malee/412725/document/DLOverview.pdf>

Note

The first version of this paper was presented at the conference “Service Tailored to the 21st Century User Needs: the Implementation of ICT in Libraries”, held on 10th December 2015 at Belgrade City Library.

Adjectival Attributes with the Nouns “čovек”, “žena”, “muškarac” and “muž”

UDC 811.163.41'322.2

DOI 10.18485/infototeca.2017.17.2.4

Marija Pantić

p.mmmmmm@gmail.com

University of Belgrade

Faculty of Philology

ABSTRACT: This research aims to produce a systematic account of the representation of men and women in the Serbian language by analysing adjectival attributes of the Serbian nouns ČOVEK (MAN), ŽENA (WOMAN/FEMALE/WIFE), MUŠKARAC (MALE) and MUŽ (HUSBAND). The analysis was performed on data obtained from the Corpus of Contemporary Serbian Language (SrpKor2013) in its advanced search mode, and it encompasses all adjectival words that occur with sufficient frequency with at least one of the four nouns. Selecting the adjectival words with an attributive function was done by reviewing the concordances, and the selected adjectival words were subjected to further analysis and classification.

KEYWORDS: adjectival attributes, adjectival words, corpus analysis, frequency lists, SrpKor2013, gender differences

PAPER SUBMITTED: 3 November 2017

PAPER ACCEPTED: 14 December 2017

1 Introduction

This paper deals with the similarities and differences in the representation of male and female persons in the Corpus of Contemporary Serbian Language through the analysis of the distribution of adjectival words occurring in the

grammatical role of adjectival attributes before the nouns ČOVEK (MAN), ŽENA (WOMAN/FEMALE/WIFE), MUŠKARAC (MALE) and MUŽ (HUSBAND).¹

Visual culture plays an important part in establishing and maintaining sexist stereotypes. The results of a Google image search for the queries “muškarac” and “žena” match a modified, not the general, meaning of the nouns that constitute these queries. The modification is explicit in the headlines of the sources of the photographs that the search returns: the noun MUŠKARAC is featured as the head of the noun phrases *neodgovoran* MUŠKARAC (*an irresponsible* MAN), *pravi* MUŠKARAC (*a real* MAN), *savršen* MUŠKARAC (*a perfect* MAN), *idealan* MUŠKARAC (*an ideal* MAN), *zreo* MUŠKARAC (*a mature* MAN) и *odgovoran* MUŠKARAC (*a responsible* MAN), while the noun ŽENA is featured as the head of the noun phrases *super* ŽENA (*a super* WOMAN), *neodoljiva* ŽENA (*an irresistible* WOMAN), *idealna* ŽENA (*an ideal* WOMAN), *savršena* ŽENA (*a perfect* WOMAN) и *najlepša* ŽENA (*the most beautiful* WOMAN). All these noun phrases are examples of modification of the meaning of the nouns by adjectives in the role of adjectival attributes. The most obvious, however, is the implicit modification based on age: the meaning of these noun phrases is illustrated mostly by photographs of men and women between the ages of twenty and forty. The age criterion may not be explicit in the accompanying text either, but the photographs themselves illustrate the noun phrases in accordance with the presumed, expected or desirable qualities of the persons they refer to (Baker, 2014, 75).

This study does not deal with implicit modification, so the dominant expectations, the issue of desirability and the presumed values are out of its scope. The author's assumption is that the perception and the conception of the meaning of nouns is greatly influenced by adjectival words that modify them (Hoey, 2004), and the aim of this paper is to present the patterns of such modification that have been discovered for the selected nouns in the Corpus of Contemporary Serbian Language. It is the author's hope that other researchers with an interest in gender and language and who are working in other disciplines might be inspired to provide various perspectives on the socioeconomic, cultural, linguistic and psychological reasons behind the data presented here, as its interpretation is beyond the scope of this paper.

¹ All English translations are by the author and based on the primary meanings of the words in their contemporary usage. The English nouns FEMALE and MALE are approximate translations, with the latter allowing a distinction to be made between the Serbian nouns ČOVEK and MUŠKARAC, which would otherwise both be translated as MAN, while the former introduces a corresponding distinction between different meanings of the noun ŽENA.

Adjectival attributes are adjectival units, that is, adjectival words (adjectives, adjectival pronouns and ordinal numbers) and adjectival phrases, the meaning of which is attributed to the nominal concept denoted by the head of the noun phrase. This attribution is formally marked by gender, number and case agreement. The syntactic and semantic differences among adjectival attributes are reflected in how tightly or loosely they are tied to the nominal concept, that is, in their position: closer to the noun or further away from it (and closer to the beginning of the phrase) (Станојчић и Поповић, 2000). This research covers adjectival words² that function as adjectival attributes and, either independently or as part of adjectival phrases, occur immediately before the nouns ŽENA, MUŠKARAC, MUŽ and ČOVEK, that is, those adjectival words that are most tightly tied to the nouns under consideration. In Serbian, adjectival attributes usually occur before the noun, so adjectival words that occur in the same function after the noun are not analysed on this occasion.

The nouns ČOVEK, ŽENA and MUŠKARAC were selected for analysis as nouns that are primarily or to a significant extent used to refer to adult female and male persons. In Matica srpska's Serbo-Croatian dictionary *Rečniku srpskohrvatskog književnog jezika* these meanings are given in definitions 2. for the noun ČOVEK and 1.6. for the noun ŽENA, while no criteria of age exist in the definitions of the noun MUŠKARAC or the definition 1.a. of the noun ŽENA. The noun ŽENA is also defined in 2.a. as “a married partner³ of the female⁴”, and in 2.6. as “the one who is married or was married”, while the meaning of the noun ČOVEK is defined in 5. as “a husband, a [male] spouse”. The meaning of the noun MUŽ is defined in 1.a. as “a married man, a male married partner, a [male] spouse” and in 1.6. as “an adult male, a man”. The usage of the noun MUŽ in the meaning 1.a. is parallel to the usage of the noun ŽENA in the meanings 2.a. and 2.6., so its inclusion in the set of the nouns to be examined allows for a partial but balanced coverage of the

² That is, those adjectival words generated by the regular expression [pos="A"], which include adjectives, ordinal numbers and adjectival pronouns. Adjectival pronouns require different search queries and a different approach to the classification of the results, so they have been omitted from this study and can be the subject of a future one.

³ The translations of dictionary definitions are adapted to English usage (for example, “bračni drug” translated as “a married friend” or “a married companion” would make little sense here).

⁴ The grammatical gender of the Serbian nouns is made explicit here for convenience.

nouns that refer to male and female spouses. Different meanings of each of the four nouns are not treated separately in this study, and this has to be kept in mind when interpreting its results. There is a certain degree of synonymy between the nouns ČOVEK, MUŠKARAC and MUŽ, and most of their meanings are antonymous with different meanings of the noun ŽENA. The most important exception is the generic meaning of the noun ČOVEK, which can partly explain the difference in the number of results returned by the corpus search for the nouns ČOVEK, MUŠKARAC and MUŽ on the one hand (87,522 results) and the noun ŽENA on the other (56,402).

2 Some related research

An exhaustive overview of the research that has influenced the approach to and the treatment of this paper's topic is not possible here, but the most important ideas and findings ought to be mentioned.

Carmen Rosa Caldas-Coulthard and Rosamund Moon have had a decisive influence on affirming a corpus approach to gender inequalities in language. Their pioneering research, presented at a symposium dedicated to critical discourse analysis, which was held at the University of Birmingham in 1999, shows that the corpus search can yield solid evidence of discrimination, sexualisation and the construction of gender in the media. They initially analysed the adjectival attributes of the nouns MAN, WOMAN and GIRL in the British press (Caldas-Coulthard and Moon, 2010), but the expanded version of the original presentation, published in 2010, includes the noun BOY as well.

Using the entire Bank of English corpus, Moon has conducted research into gender and age stereotypes and prejudices by analysing adjectival premodification of noun phrases headed by the nouns MAN/MEN and WOMAN/WOMEN modified by adjectival attributes that clearly denote age, such as *young* and *old*. Her results show that young men and women are mostly described by adjectives with positive connotations, while adjectives used to describe older men and women often have negative connotations (Moon, 2014).

Cvetana Krstev and colleagues have applied methods of corpus and computational linguistics, as well as methods of information extraction, in two studies of the representation of women in the Serbian press (Gucul-Milojević et al., 2010; Крстев и др., 2015). At the centre of their research is the identification and classification of personal names. Their results show that women are much less represented in the Serbian press than

men and that prominent women get significantly less room than prominent men.

Michael Pearce has used Sketch Engine⁵ to examine the grammatical roles of object and subject for the nouns MAN and WOMAN, as well as their adjectival attributes, in the British National Corpus.⁶ He has relied primarily on Sketch differences, that is, an automatically generated overview of similarities and differences in the grammatical and collocational behaviour of two selected words, based on which he concludes that, in the corpus analysed, the stereotypical representations of men emphasise such traits as competitiveness, adventurousness, independence, rationality, aggression, physical strength and muscularity; while the stereotypical representations of women emphasise co-operativeness, gentleness, dependency, emotionality, sympathy and physical weakness. Pearce observes that certain patterns have an uneven distribution in the BNC. For example, prose literature contains 61.45% of the noun phrases in which adjectives denoting states of neuroticism modify the noun WOMAN, even though prose literature makes up only 16% of the word count of the BNC (Pearce, 2008).

Paul Baker has researched various manifestations of sexism in the Corpus of Historical American English (COHA)⁷ and has discovered that words referring to men are more frequent than those referring to women, that nouns referring to men come first in co-ordinate noun phrases referring to men and women, and that the names of occupations with the suffix -MAN have a higher frequency than those with the suffix -WOMAN. He stresses that the noun MAN is still used with its generic meaning, despite numerous criticisms of such usage (Baker, 2014, 73–104).

The results of the application of corpus methods to the investigation of the complex and historically changeable relations between gender and language are often fascinating, and sometimes shocking, and they attract the attention of the wider public, especially when the subject of analysis is the mass media, which can show very sharp differences in the representation

⁵ A commercial corpus management and corpus query software developed by Lexical Computing Ltd, a company that was founded by Adam Kilgariff in 2003 (Kilgariff et al., 2004).

⁶ <https://goo.gl/NYgNN7> (Accessed 6 December, 2016).

⁷ COHA is composed of more than 400 million words of text in more than 100,000 individual texts published between 1810 and 2009. All the texts are categorised according to year and decade, allowing the evolution of a word or phrase, in terms of its frequency or meaning, to be examined over time. <https://goo.gl/mCxg4X> (Accessed 10 December, 2016).

of men and women. Sexist patterns that are specific to various spheres of interest can be revealed by analysis of specialised corpora such as sports news⁸ and film scenarios.⁹

3 SrpKor2013

3.1 Composition and structure

SrpKor2013 is an electronic, synchronic corpus of contemporary Serbian language that contains 122 million words. Balance in SrpKor can be expected in future, given its dynamic character and regular updating. The corpus includes 4,890 texts of the 5,058 that have been prepared for it, which have been classified into texts originally written in Serbian (4,545 texts) and translations into Serbian (513), and according to functional styles into literary (348), scientific and popular-scientific (188), newspaper (3,245), administrative (923) and other (354). The translations into Serbian cover the period from the eighteenth century to the present, while the texts written in Serbian cover the period from 1910 to 2012. With the exception of a small number of transcripts of spoken interviews, SrpKor consists of written texts, and in the process of their collection, preference was given to the already available electronic texts of various formats, while a relatively small number of non-electronic texts were digitised. Corpus texts are annotated with bibliographic information, and tokens of corpus texts are annotated with morphological information, such as part-of-speech and lemma, so that the corpus can be searched both as a collection of texts and according to information present in bibliographic and morphosyntactic annotation (Utvić, 2014, 257–259).

3.2 Alphabet

The unification of coding for the letters of the Serbian alphabets was achieved through the application of the modified version of the codes that were used in Duško Vitas's system AURORA (Vitas, 1982). Table 1 shows

⁸ University of Cambridge Research. 2016. "Aesthetics over athletics when it comes to women in sport". University of Cambridge News, 12 August 2016. <https://goo.gl/R5iLCK> (Accessed 29 December, 2016).

⁹ Hanah Anderson and Matt Daniels. 2016. "Film dialogue from 2,000 screenplays, broken down by gender and age". The Pudding, April 2016. <https://goo.gl/BbeNFS> (Accessed 23 November, 2016).

Diacritics				Digraphs			
Uppercase	AURORA code	Lowercase	AURORA code	Uppercase	AURORA code	Lowercase	AURORA code
Č	CY, Cy	č	cy	NJ, Nj	NX, Nx	nj	nx
Ć	CX, Cx	ć	cx	LJ, Lj	LX, Lx	lj	lx
Đ	DX, Dx	đ	dx	DŽ, Dž	DY, Dy	dž	dy
Š	SX, Sx	š	sx				
Ž	ZX, Zx	ž	zx				

Table 1. Serbian Latin alphabet diacritics and digraphs in AURORA code

how this code scheme maps the diacritics and digraphs used in the Serbian Latin alphabet (Utvić, 2014, 279–281).¹⁰

3.3 Search

In the process of indexing, the text of the corpus is treated not as a sequence of characters but as a sequence of token positions in the corpus, or corpus positions. This enables direct access to every token of an electronic corpus, which makes the search more efficient (Utvić, 2014, 152). All tokens of the corpus have the positional attributes **word** (a concrete realisation of a token in the text), **lemma** (assigned to a token through automatic annotation by TreeTagger), and **pos** (part-of-speech information assigned to a token through automatic annotation by TreeTagger) (Utvić, 2011).¹¹ The morphosyntactic notation applied in the morphological electronic dictionary of Serbian in the LADL/DELA format¹² was also used in the annotation of the corpus, with the tags selected restricted to the basic set, which includes ten tags for different parts of speech in Serbian and six tags for specific tokens that require special treatment.¹³

¹⁰ <https://goo.gl/5kcBjS> (Accessed 22 March, 2017).

¹¹ <https://goo.gl/cVVE4> (Accessed 22 March, 2017).

¹² The morphosyntactic description in the morphological electronic dictionary of Serbian uses the DELA formats (Dictionnaire électronique du LADL), developed by the French Laboratory for Automatic Documentation and Linguistics (Laboratoire d'automatique documentaire et linguistique, LADL) (Utvić, 2014; Krstev and Vitas, 2009).

¹³ <https://goo.gl/Bqsgih> (Accessed 22 March, 2017).

The corpus is managed by IMS Corpus Workbench,¹⁴ a collection of corpus tools, the main component of which is the query processor CQP (Corpus Query Processor). The web-based interface of SrpKor enables simple and advanced search, with the advanced search utilising most of the potential of the CQP query language CQL (CQP Query Language). The regular expression syntax that CQL is based on is a subset of the syntax of POSIX regular expressions (Utvić, 2014).

The simple search is the search of the corpus texts, not of annotation, and uses only the first positional attribute (**word**), so that it is sufficient, unlike in the advanced search, to use only the value of the positional attribute without the accompanying quotation marks, but the values of the positional attribute have to be encoded by the code AURORA. The advanced search enables a search of the morphological annotation of the corpus, that is, of the positional attributes **pos** and **lemma**. Regular expressions can be used in both the simple and advanced searches, in the first case within the value of the positional attribute **word**, and in the second as regular expressions over tokens. The corpus search generates results in the form of concordances that can be displayed in KWIC (Key Word in Context) and KWIP (Keyword in Paragraph) formats (Utvić, 2014, 308–313). The present research is based on the results generated by the advanced search and accessed in the KWIC format mostly with the default values. For a certain number of queries the results were examined page by page, as this was the only way to resolve ambiguity, polysemy and homonymy of forms.

3.4 Female authors in the corpus

Data on authorship is available for 713 texts of the 5,058 that have been prepared for the corpus, with 348 texts belonging to the literary style, 130 to the scientific and popular-scientific style, 67 to the newspaper style and 171 to other functional styles. The contribution of female authors is 52.5¹⁵ texts in the scientific and popular-scientific style (40.4%), 23 texts in the newspaper style (34.3%), 30 texts in the literary style (8.6%) and 21.33¹⁶ texts in other functional styles (12.5%).

¹⁴ <https://goo.gl/5AYJDC> (Accessed 22 March, 2017).

¹⁵ 0.5 stands for a text by two authors one of whom is female.

¹⁶ 0.33 stands for a text by three authors one of whom is female.

4 The results

4.1 The initial set

The corpus search was conducted in several phases. The initial set of nouns did not include the noun MUŽ, so the initial tests of the hypothesis about the differences in the distribution of adjectival words in the role of adjectival attributes with the nouns that denote adult male and female persons were done for the noun phrases headed by the nouns ČOVEK, MUŠKARAC and ŽENA. As discussed in the introduction, the search was restricted to the adjectival words that occur before the nouns.

Eighty-five adjectival words were selected for examination, and most of them were classified into 35 antonym pairs or pairs of sets of synonyms and their antonyms. The search was performed using the queries that specify the relevant lemmas,¹⁷ which meant that in the case of polysemous and homonymous lemmas a detailed review of the concordances was necessary to eliminate irrelevant results.

The analysis of the results has shown that two adjectival words from a particular synonym or antonym pair can have opposite frequency patterns, so that one of them occurs more frequently than the other with two of the three nouns. For example, in the antonym pair *mršav-debeo* (*thin-fat*), the adjective *debeo* has a higher frequency than its antonym with the noun ŽENA, while the reverse applies to the nouns ČOVEK and MUŠKARAC; in the pair *nag-go* (*nude-naked*), the adjective *nag* has a higher frequency than its synonym with the nouns MUŠKARAC and ŽENA, while the reverse applies to the noun ČOVEK; in the antonym pair *srećan-nesrećan* (*happy-unhappy*), the adjective *srećan* has a higher frequency than its antonym with the nouns ČOVEK and MUŠKARAC, while the reverse applies to the noun ŽENA.

However, adjectival words that were selected for this analysis were assumed to have wildly different patterns of distribution and frequencies, and even different meanings or connotations, with the nouns under consideration, and such an assumption in their selection introduces an important limitation for the generalisation of any conclusions that can be drawn from the data obtained. Objectivity could be achieved only through an exhaustive search of all the adjectival words in the relevant function and position, and by then selecting the most frequent among them for further analysis.

¹⁷ For example the search queries for the noun phrases *uspešan ČOVEK*, *uspešan MUŠKARAC* and *uspešna ŽENA* are [lemma="uspesxan"] [lemma="cyovek"], [lemma="uspesxan"] [lemma="musxkarac"] and [lemma="uspesxan"] [lemma="zxena"].

NOUN	QUERY 1	MATCHES
ČOVEK	[lemma="cyovek"]	65,345
ŽENA	[lemma="zxena"]	56,402
MUŠKARAC	[lemma="musxkarac"]	15,183
MUŽ	[lemma="muzx"]	6,994
NOUN	QUERY 2	MATCHES
ČOVEK	[pos="A"] [lemma=cyovek]	20,650
ŽENA	[pos="A"] [lemma="zxena"]	11,606
MUŠKARAC	[pos="A"] [lemma="musxkarac"]	2,890
MUŽ	[pos="A"] [lemma="muzx"]	1,129

Table 2. Frequency of the nouns in the corpus and frequency of the adjectives preceding them.

4.2 Frequency lists

In the next phase, the corpus search results generated by the search queries [pos="A"] [lemma="cyovek"], [pos="A"] [lemma="zxena"], [pos="A"] [lemma="musxkarac"] and [pos="A"] [lemma="muzx"] were normalised to the canonical forms of the adjectival words and the nouns, that is, their lemmas, and ordered by frequency using the corpus processing suite Unitex.¹⁸ The initial set of nouns is expanded by the last query to include the noun MUŽ, which, in its meaning “a married man, a male married partner, a [male] spouse”, defined in 1.a., is parallel to the very common meaning of the noun ŽENA, so it was not necessary to exclude from the search results for the search query [pos="A"] [lemma="zxena"] those in which the noun ŽENA the meaning “a married partner of the female sex, a [female] spouse”, defined in 2.a.. Additionally, the noun MUŽ also has the meaning “an adult male, a man”, defined in 1.6. (PMC, 1967-1976).

A general insight into the frequency of the nouns under consideration and the adjectival words before them was obtained using the search queries given in Table 2.

In accordance with the annotation of the corpus, query 2 does not specify the forms of the noun and the adjectival word, nor does it require number, gender and case agreement, so the results containing the noun phrases sought are not identical with but are instead a subset of the results generated by

¹⁸ <http://unitexgramlab.org> (Accessed 22 March, 2017).

this search, as they include all bigrams that consist of any adjectival word¹⁹ and one of the four nouns, regardless of their grammatical relation.

The minimum frequency threshold for these bigrams was arbitrarily set to 5 occurrences in the whole corpus,²⁰ leaving aside the different frequencies of the nouns in them. In addition to the frequency of the bigrams in the whole corpus, their frequency in relation to the frequency of the nouns is also analysed. These values are from here on referred to as the frequencies of the adjectival words (or adjectives)²¹ with the nouns.²² According to the frequency lists, the minimum frequency threshold was satisfied by 335 of the bigrams containing the noun ČOVEK, 270 of those containing the noun ŽENA, 101 of those containing the noun MUŠKARAC and 22 of those containing the noun MUŽ, which gave a total of 474 adjectival words.

4.3 The review of the concordances and the classification of the results

Further analysis required that the concordances be generated for the purpose of determining if the bigrams in the frequency lists correspond to the noun phrases sought. The queries that specify the lemma of one of the 474 adjectival words followed by the lemma of one of the four nouns were used for this purpose (see footnote 17 for an example). Thus, the frequency lists served as an initial step in the selection of adjectival words, while the analysis of the concordances allowed for further examination of the morphological, syntactic and lexical-semantic properties and relations of the selected adjectival words and the four nouns under consideration. Partial incongruity of the frequency lists and the results generated by the corpus search is inevitable given the methodological differences between the two approaches. Those adjectival words that did not occur in a sufficient number of relevant results generated by the corpus search were excluded from further analysis.

¹⁹ That is, any adjectival word generated by the regular expression [pos="A"].

²⁰ Frequencies lower than 5 instances in the whole corpus were equated with 0.

²¹ For an explanation of the difference between the two terms, see the introduction and footnote 1. The latter (narrower) term will be used wherever adjectives alone are discussed (not including participles).

²² For example, the noun ŽENA occurs 56,402 times in the corpus, and the noun ČOVEK 65,345 times. An adjective that occurs 100 times with each of the nouns will have the frequency of 0.18% with the noun ŽENA and of 0.15% with the noun ČOVEK. Therefore, its frequency is higher with the noun ŽENA.

The analysis of the concordances has also shown that certain adjectival words, such as *divalj* (*wild*), *izgubljen* (*lost*), *mali* (*small*), *slobodan* (*free*) and *velik* (*big*), can have different meanings depending on which of the four nouns they modify. These adjectival words were assigned to a group of their own, the analysis of which is not presented in this paper.

All the other adjectival words are classified into the following groups:

1. Age;
2. Physical appearance;
3. State;
4. Predispositions;
5. Subjective evaluation;
6. Social and personal status;
7. The personality model;
8. Emphasis and distinction;
9. Belonging;
10. Relations.

All these groups were formed as lexical fields based on the results from the corpus specifically for the purpose of this research, and the selection of their members depended primarily on the author's intuition. They may partially overlap with different semantic, lexicological and computational models, but they are not based on them. Group 7 was formed in accordance with the five-factor personality model (Docs, 2017), which, considering its nature and purpose, could not supply the members of this group, so, given the approach to their selection, this group is not fundamentally different from the other groups. This classification is intended to facilitate a clearer representation of the data, but in interpreting the results for each group it is necessary to keep in mind which adjectival words it includes, given the inevitable subjectivity of their selection.

All adjectival words included in these groups have occurred in the corpus search with a meaning that corresponds to a given group and the function of an adjectival attribute in at least 5 results with at least one of the nouns. The queries for some of them, however, had generated the results that were not taken into consideration in cases of irrelevant functions, low frequencies of other meanings for polysemous adjectival words and wrong parts of speech for polysemous and homonymous lemmas. For example, the adjective *pun* (*full*) was included in the meaning 6.a. "the one who is moderately fat, fattened, large, plump" (PMC, 1967-1976), but it occurs with other meanings

in an insufficient number of results (example 1) some of which do not correspond to the function being sought (example 2). The query [lemma="pun"] [lemma="cyovek"] returns the adverb *puno* as well (example 3).

example 1 ...*da bi ispunio misiju pravog i punog čoveka* (...to fulfil the mission of the real and complete man.).

example 2 *Posramljeni u čekaonici punoj žena...* (Embarrassed in the waiting room full [of] women)

example 3 ...*kod nas ima puno žena slikara...*(...we have many women painters...)

Group 1: Age. This group includes the following adjectives:

Serbian	English
<i>mlad, star,²³ sredovečan, postariji, tridesetogodišnji, dvadesetpetogodišnji, četrdesetogodišnji, pedesetogodišnji, odrastao, punoletan, mator, ostareo, vremešan, zreo, godišnji.</i>	<i>young, old, middle-aged, elderly, thirty-year-old, twenty-five-year-old, forty-year-old, fifty-year-old, grown-up/adult, adult, age-old, aged, timeworn, mature, year-old.</i>

Adjectives belonging to group 1 (chart 1 in the Appendix) occur with all four nouns and have the highest frequency with the noun ŽENA (3%). In this group, the adjective that has the highest frequency with one of the nouns

²³ The Serbian adjective *star* is here included in the meaning 1.a. “who has lived for many years”, but it can also denote a belonging to a sequence (group 9, subgroup 9.1), in its meanings 2.a. “the one that was in the past, former” and 2.6. “previous, earlier, past”, but also a belonging to a culture (group 9, subgroup 9.2) in the meaning 4. “who, in characteristics or outlook, belongs to a previous time” and a social status derived from experience, importance or rank (group 6, subgroup 6.4) in the meaning 6.a. “long-standing, long known for something” and 8.b. “higher in rank, higher in some hierarchy”. The search query [lemma="star"] [lemma="cyovek"] " generates less than 3% of results in which the adjective *star* has one of these meanings, while the search query [lemma="star"] [lemma="zxena"] returns only one result (out of 317) that does not have the meaning defined in 1.a. The search query [lemma="star"] [lemma="muzx"] also returns one result with the meaning 2.6., and additionally one result in which the meaning 1.a. is not necessarily the primary one. With the noun MUŠKARAC the adjective *star* occurs exclusively in the meaning 1.a. (PMC, 1967-1976).

is the adjective *mlad* with the noun ČOVEK (2.28%), and the most frequent noun phrases are *mlad* ČOVEK (1,488), *mlada* ŽENA (1,196) and *stara* ŽENA (317).

Group 2: Physical appearance. Depending on what aspect of physical appearance they describe, these adjectival words are classified into three subgroups:

Subgroup	Serbian	English
2.1. Colour	<i>plavokos, beo, crn, plav, plavook, sed, žut, bled, crnomanjast, tamnoput;</i>	<i>blond(e)-haired, white, black, blond(e), blue-eyed, grey, yellow, pale, dark, dark-skinned;</i>
2.2. Figure	<i>visok, gojazan, punačak, vitak, debeo, mršav, krupan, nizak, onizak, ogroman, sitan, pun, razvijen;</i>	<i>tall, obese, chubby, slender, fat, thin, large, short, shortish, huge, small, full, muscular</i>
2.3. Current appearance	<i>nag, go, odeven, razgoličen, obučen,²⁴ zabrađen, doteran, naoružan, brkat, bradat, maskiran.</i>	<i>nude, naked, clothed, scantily dressed, dressed, head-scarved, well-groomed, armed, moustached, bearded, masked.</i>

Adjectival words belonging to group 2 (chart 2) do not occur with the noun MUŽ, and they have the highest frequency with the noun MUŠKARAC (1.4%). In this group, the adjectival word that has the highest frequency with one of the nouns is the participle *naoružan* with the noun MUŠKARAC (0.4%), while the most frequent noun phrases are *beo* ČOVEK (81), *visok* ČOVEK (76) and *naoružan* MUŠKARAC (61). The most frequent with nouns ČOVEK and ŽENA are adjectives of the subgroup 2.1, and with the noun MUŠKARAC adjectival words of the subgroup 2.3.

Group 3: State. Depending on the type of state that they describe, these adjectival words are classified into three subgroups.

²⁴ This is the passive participle of the verb *obučiti* (*to dress*), which is homographic with the passive participle of the verb *obučiti* (*to train*) in the corpus search, because of its lack of accents. The passive participle of the verb *obučiti* occurs once, with the noun ŽENA, and that result is not included here.

Subgroup	Serbian	English
3.1. Emotional	<i>srećan, nesrećan/nesretan, unesrećen, zaljubljen, razočaran, očajan, tužan, uplašen, uznemiren;</i>	<i>happy, unhappy, grief-stricken, in love, disappointed, desperate, sad, scared, upset;</i>
3.2. Permanent physical	<i>živ, mrtav/umrli, pokojni, preminuli, smrtan;</i>	<i>alive, dead, deceased, departed, mortal;</i>
3.3. Current psycho-physiological	<i>bolestan/oboleo, zaražen/inficiran, zdrav, gladan, umoran, trudan, noseći, bremenit, pijan, usamljen, uplakan, preporođen.</i>	<i>sick/ill, infected, healthy, hungry, tired, pregnant, carrying, bearing, drunk, lonely, crying/tearful, rejuvenated.</i>

Adjectival words belonging to group 3 (chart 3) occur with all four nouns and have the highest frequency with the noun ČOVEK (1.29%). In this group, the adjectival word that has the highest frequency with one of the nouns is the adjective *nesrećan* with the noun ŽENA (0.27%), and the most frequent noun phrases are *nesrećna ŽENA* (155), *srećan ČOVEK* (154) and *nesrećan ČOVEK* (126). The most frequent with the noun ČOVEK are adjectival words of the subgroup 3.1, with the noun ŽENA adjectival words of the subgroup 3.3, and with the nouns MUŠKARAC and MUŽ adjectival words of the subgroup 3.2.

Group 4: Predisposition. Depending on the type of predisposition that they denote, these adjectives are classified into two subgroups:

Subgroup	Serbian	English
4.1. Physical and psycho-physiological	<i>plodan, sterilan, levoruk, slep, jak/snažan, brz, slab, heteroseksualan, sposoban, feminiziran;</i>	<i>fertile, sterile, left-handed, blind, strong, fast, weak, heterosexual, capable, effeminate;</i>
4.2. Intellectual	<i>genijalan, inteligentan, talentovan, darovit, pametan, glup.</i>	<i>genius, intelligent, talented, gifted, smart, stupid.</i>

Adjectives belonging to group 4 (chart 4) do not occur with the noun MUŽ, and they have the highest frequency with the noun ČOVEK (0.83%). In this group, the adjective that has the highest frequency with one of the nouns is the adjective *pametan* with the noun ČOVEK (0.25%), and the most frequent noun phrases are *pametan* ČOVEK (166), *brz* ČOVEK (77) and *jak* ČOVEK (68). The most frequent with the nouns ČOVEK and MUŠKARAC are adjectives of the subgroup 4.1, and with the noun ŽENA adjectives of the subgroup 4.2.

Group 5: Subjective evaluation. Group 5 includes adjectives that convey a subjective evaluation, and, depending on what is being evaluated, they are classified into four subgroups:

Subgroup	Serbian	English
5.1. Subjective evaluation of effect	<i>privlačan/atraktivnan, poželjan, fatalan, zanosan, prijatan, strašan, odvratnan, zagonetan, šarmantan, opasan, drag, mio, čudan, čudnovat, dražestan, duhovit, zanimljiv, tajanstven, simpatičan, ljubak, misteriozan, mračan, otmen, prost/jednostavan, komplikovan, nezgodan;</i>	<i>attractive, desirable, fatal,²⁵ captivating, pleasant, frightening, disgusting, intriguing, charming, dangerous, dear, sweet, weird, peculiar, delightful, witty, interesting, secretive, appealing, lovely, mysterious, gloomy, decorous, simple, complicated, difficult;</i>
5.2. Subjective evaluation of appearance	<i>lep, prelep, ružan, elegantan, naočit, stasit, zgodan;</i>	<i>beautiful, gorgeous, ugly, elegant, eye-catching, well-built, shapely/handsome;</i>
5.3. Subjective evaluation of morals	<i>zao, dobar/valjan, loš/rđav, ispravan, grešan, moralan, svet, lak, pao;</i>	<i>evil, good, bad, upright, sinful, moral, saintly, easy, fallen;</i>
5.4. General subjective evaluation	<i>krasan/divan/predivan, idealan, savršen, sjajan, predivan, odličan.</i>	<i>great/admirable/wonderful, ideal, perfect, brilliant, great.</i>

Adjectives belonging to group 5 (chart 5) occur with all four nouns and have the highest frequency with the noun ŽENA (1.82%). In this group, the adjective that has the highest frequency with one of the nouns is the adjective *lepa* with the noun ŽENA (0.86%), and the most frequent noun phrases are

²⁵ Only in the sense of “femme fatale”.

lepa ŽENA (487), *dobar* ČOVEK (366) and *dobra* ŽENA (78). The most frequent with the nouns ŽENA and MUŠKARAC are adjectives of the subgroup 5.2, and with the nouns ČOVEK and MUŽ adjectives of the subgroup 5.3.

Group 6: Social and personal status. Depending on the type of status that they denote, these adjectival words are classified into seven subgroups:

Subgroup	Serbian	English
6.1. Marital status	<i>udat/oženjen, bivši, budući, neudat/neoženjen, zakonit, razveden, venčan, nevenčan, sadašnji;</i>	<i>married, former, future, unmarried, lawful, divorced, wedded, unwedded, present;</i>
6.2. Educational and professional status	<i>zaposlen, nezaposlen, visokoobrazovan, obrazovan, neobrazovan, učen, školovan, stručan, nepismen, pismen, vešt;</i>	<i>employed, unemployed, highly educated, educated, uneducated, learned, schooled, professional, illiterate, literate, skillful;</i>
6.3. Financial status	<i>bogat, siromašan, imućan, prebogat;</i>	<i>rich, poor, wealthy, super-rich;</i>
6.4. General social or personal status	<i>biviši,²⁶ uspešan, moćan, nemoćan, uticajan, poznat, slavan, anonimn, znamenit, značajan, važan, ostvaren, ugledan, nevin/nedužan, nevidljiv, zaslužan, popularan, glavni/čelni, vodeći, uzoran, pobunjen, bespomoćan, jadan, sirot, iskusan,²⁷ ključni;</i>	<i>former, successful, powerful, powerless, influential, known, famous, anonymous, notable, significant, important, fulfilled, reputable, innocent/guiltless, invisible, deserving, popular, main/head, leading, exemplary, rebellious, helpless, miserable, unfortunate, experienced, key;</i>
6.5 Emotional status	<i>omiljen, voljen, ljubljen</i>	<i>favourite, loved, beloved;</i>

²⁶ The noun phrase *biviši* ČOVEK is an expression with the meaning “ruined man” (PMC, 1967-1976).

²⁷ The adjective *iskusan* with the noun ČOVEK has the meaning of professional and life experience, while with the noun ŽENA it most frequently refers to sexual experience.

6.6. Victimhood and criminality	<i>ubijen, silovan, zlostavljan, ranjen, pretučen, otet, vezan, prevaren, poražen, povređen, ugrožen, zarobljen, uhapšen, osuđen;</i>	<i>killed / murdered, raped, abused, wounded, beaten, kidnapped / abducted, tied, deceived / cheated, defeated, injured, endangered, imprisoned, arrested, convicted;</i>
6.7. Definition and a change of status	<i>pregledan, ispitan, anketiran, izložen, postavljen, tražen, neimenovan, neidentifikovan, izabran.</i>	<i>examined, questioned, surveyed, exposed, appointed, wanted, unnamed, unidentified, chosen / selected.</i>

Adjectival words belonging to group 6 (chart 6) occur with all four nouns and have the highest frequency with the noun ČOVEK (2.38%). In this group, the adjectival word that has the highest frequency with one of the nouns is the adjective *bivši* with the noun MUŽ (1.82%), and the most frequent noun phrases are *bogat ČOVEK* (238), *glavni ČOVEK* (155) and *čelni ČOVEK* (149). The most frequent with all four nouns are adjectival words of the subgroup 6.4.

Group 7: Personality model These adjectival words are classified according to the five-factor model of personality into the following five subgroups:

Subgroup	Serbian	English
7.1. Emotional stability	<i>depresivan, frigidan, miran, neurotični, obezglavljen, smiren, poremećen, zadovoljan, histeričan, ljubomoran, vatren, normalan, lud, osetljiv;</i>	<i>depressed, frigid, peaceful, neurotic, frantic, calm, disturbed, satisfied, hysterical, jealous, fiery, normal, crazy, sensitive;</i>
7.2. Extroversion	<i>aktivan, otresit, odlučan, vedar, nasmejan, društven, energičan, raspoložen, veseo, neustrašiv, povučen, privatn, tih, ćutljiv, zatvoren, stidljiv, preduzimljiv, ambiciozan;</i>	<i>active, bold, resolute, bright, smiling, sociable, energetic, cheerful, merry, fearless, withdrawn, private, quiet, silent, introverted, shy, enterprising, ambitious;</i>

7.3. Agreeableness	<i>častoljubiv, ljubazan, pristojan, pažljiv, kulturn, iskren, poverljiv, sujetan, samoljubiv, dobronameran, pošten, čestit / častan, vaspitan, plemenit, nasilan, agresivan, human, pravedan, nepravičan, surov, skroman, pravičan, tolerantan, prirodan, lukav, fin;</i>	<i>fame-seeking, kind, decent, attentive, cultured, sincere, confidential, vain, self-loving, well-meaning, honest, honourable, [well] bred, noble, violent, aggressive, humane, just, unfair, cruel, modest, fair, tolerant, natural, cunning, nice;</i>
7.4. Openness	<i>kreativan, strastan, svestran, hrabar, odvažan, samosvestan, prosvetćen, emancipovan, blag, pragmatičan, praktičan, neupućen, nezavisan, naivan, zainteresovan;</i>	<i>creative, passionate, versatile, brave, bold, self-aware, enlightened, emancipated, mild, pragmatic, practical, uninformed, independent, naïve, interested;</i>
7.5 Conscientiousness	<i>umeren, odmeren, pouzdan, vredan / radan, odgovoran, razborit, realan, razuman, racionalan, mudar, misaon / misleći / uman, ozbiljan, civilizovan, principijelan, uporan, veran, neveran, porodičan, poslovan, dostojanstven, neodgovoran, lenj, oprezan, trezven, strog.</i>	<i>moderate, measured, reliable, industrious / diligent, responsible, sensible, realistic, reasonable, rational, wise, thoughtful / thinking, serious, civilised, principled, persistent, faithful, unfaithful, family, business, dignified, irresponsible, lazy, cautious, sober, strict.</i>

Adjectival words belonging to group 7 (chart 7) occur with all four nouns and have the highest frequency with the noun ČOVEK (3.62%). In this group, the adjectival word that has the highest frequency with one of the nouns is the adjective *poslovan* with the noun ČOVEK (0.48%), and the most frequent noun phrases are *poslovan* ČOVEK (313), *poslovna* ŽENA (169) and *pošten* ČOVEK (163). The most frequent with the nouns ČOVEK and ŽENA are adjectival words of the subgroup 7.5, with the noun MUŠKARAC adjectival words of the subgroups 7.5 and 7.3, and with the noun MUŽ adjectival words of the subgroup 7.1.

Group 8: Emphasis and distinction. Adjectives belonging to group 8 (chart 1) determine the extent to which a meaning of the noun that is common or dominant in a given social or personal context applies to the person the noun refers to.

Serbian	English
<i>prav, stvaran, istinski, poseban, izuzetan, prosečan, neponovljiv, običan, neobičan, redak, drugačiji, drukčiji, jedinstven, izvanredan.</i>	<i>real, actual, true, special, exceptional, average, unrepeatable, ordinary, unusual, rare, different, unique, extraordinary.</i>

These adjectives do not occur with the noun MUŽ, and they have the highest frequency with the noun ČOVEK (1.25%). In this, group, the adjective that has the highest frequency with one of the nouns is the adjective *običan* with the noun ČOVEK (0.69%), and the most frequent noun phrases are *običan ČOVEK* (448), *pravi ČOVEK* (125) and *prosečan ČOVEK* (63).

Group 9: Belonging. Depending on the type of belonging that they attribute to the person the noun refers to, these adjectival words are classified into seven subgroups:

Subgroup	Serbian	English
9.1. Sequence	<i>prvi, drugi, treći, četvrti, peti, šesti, sedmi, deseti, dvanaesti, poslednji;</i>	<i>the first, the second, the third, the fourth, the fifth, the sixth, the seventh, the tenth, the twelfth, the last;</i>
9.2. Period and culture	<i>savremen, moderan, današnji, drevan/prvobitan, praistorijski, renesansni, antički, srednjovekovan, pečinski, starozavetni, primitivan;</i>	<i>contemporary, modern, today's, primeval/primordial, prehistoric, Renaissance, ancient,²⁸ medieval, cave, Old Testament, primitive;</i>
9.3. Place and culture	<i>ruski, zapadni, evropski, svetski, američki, gradski, srpski, lokalni/ovdašnji/tamošnji, sovjetski, šidski, narodski/narodni, urbani, seoski, turski, romski, albanski, balkanski, arapski, avganistanski, iranski, španski, bošnjački, britanski, nemački;</i>	<i>Russian, Western, European, world, American, city, Serbian, local/[of this place]/[of that place], Soviet, Šid's, people's, urban, rural/village, Turkish, Romani, Albanian, Balkan, Arab, Afghan, Iranian, Spanish, Bosniak, British, German;</i>

²⁸ Only with the sense of classical antiquity

9.4. Religion	<i>religiozan, pobožan, božji, crkven, verujući, duhovan, pravoslavan, muslimanski, jevrejski;</i>	<i>religious, devout, God's, Church's, believing, spiritual, Orthodox [Christian], Muslim, Jewish;</i>
9.5. Profession	<i>fudbalski, pozorišni, javan;</i> ²⁹	<i>[of] football, [of] theatre, public;</i> ³⁰
9.6. Political and state structures	<i>Miloševićev, partijski, oligarhijski, demokratski, Dinkićev, Tadićev, knežev, carski, politički;</i>	<i>Milošević's, party, oligarchic, democratic, Dinkić's, Tadić's, prince's, imperial, political;</i>
9.7. Family and partner	<i>Petrov, oficirski, Savin, lekarev, pokojnikov, rođen / vlastit / sopstven.</i>	<i>Petar's, officer's, Sava's, doctor's, the deceased's, own.</i>

Adjectival words belonging to group 9 (chart 8) occur with all four nouns and have the highest frequency is with the noun ČOVEK (8.95%). In this group, the adjectival word that has the highest frequency with one of the nouns is the ordinal number *prvi* with the noun ČOVEK (5.66%), and the most frequent noun phrases are *prvi* ČOVEK (3,696), *prva* ŽENA (670) and *drugi* ČOVEK (696). The most frequent with all four nouns are the adjectival words of the subgroup 9.1.

Group 10: Relations. According to the type of relation that they denote, these adjectival words are classified into three subgroups:

Subgroup	Serbian	English
10.1. Personal	<i>tud, nepoznat, stran, pogrešan, blizak;</i>	<i>foreign/another's, unknown/unfamiliar, unfamiliar/strange, wrong, close;</i>

²⁹ It is important to note here that in the corpus the noun phrase *javna* ŽENA occurs as an expression with the meaning “a woman of easy morals, a prostitute”, while the adjective *javan* with the noun ČOVEK has the meaning 3. “who is associated with the sociopolitical life” and the noun phrase *javni* ČOVEK refers to a politician (PMC, 1967-1976).

³⁰ **public woman** (accessed 15 November 2017).

10.2. General	<i>konkretan, određen, pojedini, ijedan, nijedan, jedini, pojedinačan, nov, isti, brojni, ostali, mnogi, različit, prisutan</i>	<i>concrete, specific, individual, any one, not one, the only, singular, new, the same, numerous, other, many, various, present;</i>
10.3. Aspect or part	<i>spoljašnji, unutrašnji, ceo, telesni.</i>	<i>external, internal, whole, bodily.</i>

Adjectival words belonging to group 10 (chart 9) occur with all four nouns and have the highest frequency with the noun MUŠKARAC (2%). In this group, the adjectival word that has the highest frequency with one of the nouns is the adjective *nepoznat* with the noun MUŠKARAC (0.82%), and the most frequent noun phrases are *mnoge ŽENE* (208), *jedina ŽENA* (178) and *jedini ČOVEK* (175). The most frequent with all four nouns are adjectival words of the subgroup 10.2.

4.4 Overview of the results

The noun ČOVEK is most likely to be modified by the adjectival words that have been analysed, with their frequency with it in the relevant function being 25.47%, compared to 15.86% with the noun ŽENA, 13.32% with the noun MUŠKARAC and 8.62% with the noun MUŽ. (chart 10).

The noun MUŠKARAC is most frequently modified by adjectives denoting age (group 1), the nouns ČOVEK and ŽENA by adjectival words attributing some type of belonging to the person the noun refers to (group 9), and the noun MUŽ by adjectival words that attribute a certain social status to the person the noun refers to (group 6).

With the noun ŽENA the frequency of adjectives denoting age is nine times higher than with the noun MUŽ and by 3–4% higher than with the nouns ČOVEK and MUŠKARAC. Adjectival words that define an aspect of the physical appearance of the person the noun refers to (group 2) do not occur with the noun MUŽ, while their frequency with the noun MUŠKARAC is higher by about 65% than with the noun ČOVEK and almost two times higher than with the noun ŽENA. With the noun ČOVEK, the frequency of adjectival words denoting a state of the person the noun refers to (group 3) is four times higher than with the noun MUŠKARAC and by about 10% higher than with the noun ŽENA. Adjectives that denote a predisposition of the person the noun refers to (group 4) do not occur with the noun MUŽ, and their frequency with the noun ČOVEK is two-and-a-half to three times higher than with the nouns ŽENA and MUŠKARAC. With the noun ŽENA, the frequency of adjectives conveying subjective evaluation (group 5) is by about 2% higher than with the noun ČOVEK and three to three-and-a-half times higher than with the nouns MUŠKARAC

and MUŽ. With the noun MUŽ, the frequency of adjectival words that attribute a certain social status to the person the noun refers to (group 6) is higher by about 25% than with the noun ŽENA, by about 50% than with the noun ČOVEK, and it is twice as high as with the noun MUŠKARAC. With the noun ČOVEK, the frequency of adjectival words related to the five-factor model of personality is three times higher than with the nouns ŽENA and MUŠKARAC and four-and-a-half times higher than with the noun MUŽ. Adjectival words that determine the degree of conventionality of the person the noun refers to (group 8) do not occur with the noun MUŽ, and their frequency with the noun ČOVEK is five times higher than with the noun ŽENA and twice as high as with the noun MUŠKARAC. With the noun ČOVEK the frequency of adjectival words that attribute some type of belonging to the person the noun refers to (group 9) is about four-and-a-half times higher than with the noun MUŽ, almost four times higher than with the noun MUŠKARAC and almost three times higher than with the noun ŽENA. With the noun MUŠKARAC, the frequency of adjectival words that refer to an aspect or a part of the person the noun refers to, or to the relation of that person and some other persons (group 10), is almost four times higher than with the noun MUŽ, 40% higher than with the noun ČOVEK and about 55% higher than with the noun ŽENA.

To sum up, adjectival words belonging to the groups 3, 4, 7, 8 and 9 have the highest frequency with the noun ČOVEK, while the frequency of adjectival words belonging to the groups 1 and 5 is the highest with the noun ŽENA, of those belonging to the groups 2 and 10 with the noun MUŠKARAC, and of those belonging to the group 6 with the noun MUŽ.

ČOVEK		ŽENA	
Serbian	English	Serbian	English
<i>prvi</i> (3631)	<i>first</i> (3631)	<i>mlad</i> (1196)	<i>young</i> (1196)
<i>mlad</i> (1488)	<i>young</i> (1488)	<i>prvi</i> (635)	<i>first</i> (635)
<i>drugi</i> (562)	<i>second</i> (562)	<i>lep</i> (487)	<i>beautiful</i> (487)
<i>običan</i> (488)	<i>ordinary</i> (488)	<i>drugi</i> (392)	<i>second</i> (392)
<i>dobar</i> (366)	<i>good</i> (366)	<i>star</i> (317)	<i>old</i> (317)
<i>poslovan</i> (313)	<i>business</i> (313)	<i>mnogi</i> (188)	<i>many</i> (188)
<i>savremen</i> (267)	<i>contemporary</i> (267)	<i>poslovan</i> (169)	<i>business</i> (169)
<i>star</i> (262)	<i>old</i> (262)	<i>jedini</i> (159)	<i>only</i> (159)
<i>pametn</i> (166)	<i>smart</i> (166)	<i>nesrećan</i> (155)	<i>unhappy</i> (155)
<i>pošten</i> (163)	<i>honest</i> (163)	<i>udat</i> (138)	<i>married</i> (138)
		<i>uspešan</i> (138)	<i>successful</i> (138)

MUŠKARAC		MUŽ	
Serbian	English	Serbian	English
<i>mlad</i> (180)	<i>young</i> (180)	<i>bivši</i> (127)	<i>former</i> (127)
<i>nepoznat</i> (124)	<i>unknown</i> (124)	<i>pokojni</i> (66)	<i>deceased</i> (66)
<i>stran</i> (119)	<i>strange</i> (119)	<i>prvi</i> (50)	<i>first</i> (50)
<i>star</i> (118)	<i>old</i> (118)	<i>drugi</i> (38)	<i>second</i> (38)
<i>drugi</i> (114)	<i>second</i> (114)	<i>dobar</i> (27)	<i>good</i> (27)
<i>mnogi</i> (84)	<i>many</i> (84)	<i>ljubomoran</i> (24)	<i>jealous</i> (24)
<i>godišnji</i> (69)	<i>yearly</i> (69)	<i>nasilan</i> (16)	<i>violent</i> (16)
<i>naoružan</i> (61)	<i>armed</i> (61)	<i>slavan</i> (16)	<i>famous</i> (16)
<i>odrasao</i> (60)	<i>adult</i> (60)	<i>nov</i> (15)	<i>new</i> (15)
<i>oženjen</i> (50)	<i>married</i> (50)	<i>prevaren</i> (14)	<i>cheated</i> (14)
		<i>bogat</i> (14)	<i>rich</i> (14)

5 Further work

The four nouns examined in this study occur in the corpus with different meanings that can be taken into consideration in the analysis of the results. A complementary analysis of other nouns denoting male and female persons of different ages could be very informative. Different meanings and different connotations that certain adjectival words can have depending on the noun they modify also require further analysis. Analyses of the distribution of the degrees of adjectives and of antonym pairs would produce a more precise picture of the intensity and the frequency of affirmative and negative descriptions with each of the nouns, and reveal which adjectives are most often used in relative descriptions and with which nouns. For example, the adjective *star* (*old*) never occurs in its positive degree with the noun MUŠKARAC, which it modifies more frequently than the other three nouns.

A more detailed analysis of possessive adjectives would require a careful reading of the concordances and their classification based on the types of belonging and the relationships that these adjectives denote with a given noun. The same is true of possessive pronouns, which require a different search as well.

If adjectival words influence the general perception of the meaning of the nouns that they modify, it makes sense to assume that the reverse applies too, and so the frequency of the noun phrases examined here should be analysed in relation to the frequency of the adjectival words within them.

The results from SrpKor can be compared with the results of corresponding searches in the comparable corpora of other languages. As for other potential corpora of Serbian, a significantly different approach would be necessary for work with untagged corpora, but particularly interesting for historical and diachronic

insights into the representation of men and women would be the proverbs collected by Vuk Stefanović Karadžić³¹ and Matica srpska's dictionary *Rečnik srpskohrvatskog književnog jezika*, as they are products of collective consciousness at a time of ideological change and mass emancipation. The potential of dictionary analysis for studies of language and gender was demonstrated by Fournier and Russell in an interesting research into gender stereotypes in the Oxford English Dictionary (Fournier and Russell, 2004). They have shown that gender stereotypes are very frequent in citations but that they also occur in definitions, thus revealing patterns of both historical and contemporary sexism. Even on a brief examination, the aforementioned dictionary of the Serbo-Croatian language offers great inspiration for a similar undertaking. For example, for the meaning 1.a. of the noun MUŽ (HUSBAND), the definition of which is “a married man, a male married partner, a [male] spouse”, the following citation is given: “He is a husband and therefore has no work at home”. The citation, “We women are born destined to listen to men, even if they are stupid”, illustrates the meaning 2. of the conjunction AKO (EVEN IF, ALTHOUGH, THOUGH). A number of entries, such as those for the nouns ŽENSKAĆ (FEMALE), MUŽ (SCAMP) and ANATEMA (ANATHEMA), feature citations that strongly state a preference for male children or include explicit references to the undesirability of female children and to the despair of their parents. The noun AŽDAJA (DRAGON) has a derogatory meaning of “an evil, horrible woman”, defined in 2.6., and the basic meaning of “a young female horse” for the noun AJGIRUŠA is complemented by a figurative and pejorative one of “a lustful, voluptuous woman”. The verb ŠEFOVATI (TO BOSS, TO BOSS AROUND) is accompanied by the following citation: “Women work—men boss around”. The wry humour of the lexicographer is clear in some of these examples, but the choice of the citations that express a preference for male children is harder to interpret. The insight into historical attitudes through the prism of the lexicographer that this dictionary provides makes it a valuable source of social commentary and invites a comparison between past and present sensibilities.

References

- Baker, Paul. *Using corpora to analyze gender*. Bloomsbury Academic, 2014.
- Caldas-Coulthard, Carmen Rosa and Rosamund Moon. “‘Curvy, hunky, kinky’: Using corpora as tools for critical analysis”. *Discourse & Society* Vol. 21, no. 2 (2010): 99–133.
- Docs, IBM Bluemix. *Personality models*. IBM Bluemix, 2017. Accessed December 23, 2016, <https://goo.gl/MZptnU>

³¹ Compiled into an untagged corpus by Cvetana Krstev: <https://goo.gl/K6YtjU> / (accessed 13 August 2017).

- Fournier, Hannah S and Delbert W Russell. "A study of sex-role stereotyping in the Oxford English Dictionary 2E". *Computers and the Humanities* Vol. 26, no. 1 (2004): 13–20.
- Gucul-Milojević, Sandra, Vanja Radulović and Cvetana Krstev. "A view on the representation of women in Serbian newspaper texts". In *Applications of Finite-State Language Processing: Selected Papers from the 2008 International NooJ Conference*, Váradi, Tamás, Judit Kuti and Max Silberstein, 160–170. Cambridge Scholars Publishing, 2010. Accessed July 5, 2016, <https://goo.gl/42EPvA>
- Hoey, Michael. *Lexical priming: A new theory of words and language*. Routledge, 2004.
- Kilgarrriff, Adam, Pavel Rychly, Pavel Smrz and David Tugwell. "The Sketch Engine". In *Proceedings of the Eleventh EURALEX International Congress, EURALEX 2004, Lorient, France, July 6-10, 2004*, 105–115. 2004. Accessed November 17, 2016, <https://goo.gl/YEHQur>
- Krstev, Cvetana and Duško Vitas. "An effective method for developing a comprehensive morphological e-dictionary of compounds"/ In *Arena Romanistica – The 28th Conference on Lexis and Grammar, Bergen, 29th September–3rd October 2009*, 204–212. University of Bergen, Department of Foreign Languages, 2009.
- Moon, Rosamund. "From gorgeous to grumpy: adjectives, age and gender.". *Gender & Language* Vol. 8, no. 1 (2014).
- Pearce, Michael. "Investigating the collocational behaviour of MAN and WOMAN in the BNC using Sketch Engine". *Corpora* Vol. 3, no. 1 (2008): 1–29. Accessed November 15, 2016, <https://goo.gl/mn7EG8>
- Utvić, Miloš. "Izgradnja referentnog korpusa savremenog srpskog jezika". Doktorska disertacija, Univerzitet u Beogradu, Filološki fakultet, 2014. Accessed October 17, 2016, <https://goo.gl/ug1akT>
- Utvić, Miloš. "Annotating the Corpus of Contemporary Serbian". *INFOtheca* Vol. 12, no. 2 (2011): 36a–47a.
- Vitas, Duško. "Prikaz jednog sistema za automatsku obradu teksta". In *ROLP – Računarska obrada lingvističkih podataka, Bled, 7-9. X 1982.*, Tancig, Peter and Duško Vitas, Institut "Jožef Stefan", 457–461. 1982.
- Крстев, Цветана, Милош Утвић и Јелена Јаћимовић. "Ако коза лаже, рог не лаже: Где су и ко су жене у српској дневној штампи". *Књижевство* Vol. 5, (2015). Accessed July 3, 2016, <https://goo.gl/zRhFPu>
- РМС. *Речник српскохрватскога књижевног језика (I–III Матица српска – Матица хрватска, IV–VI Матица српска)*. Нови Сад: Матица српска, 1967–1976.
- Станојчић, Живојин и Љубомир Поповић. *Граматица српскога језика*. Завод за уџбенике и наставна средства, 2000.

Appendix

Charts

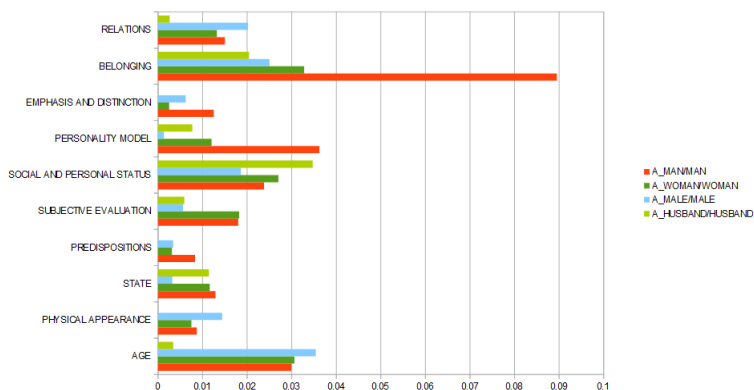


Figure 1. Overview of the groups

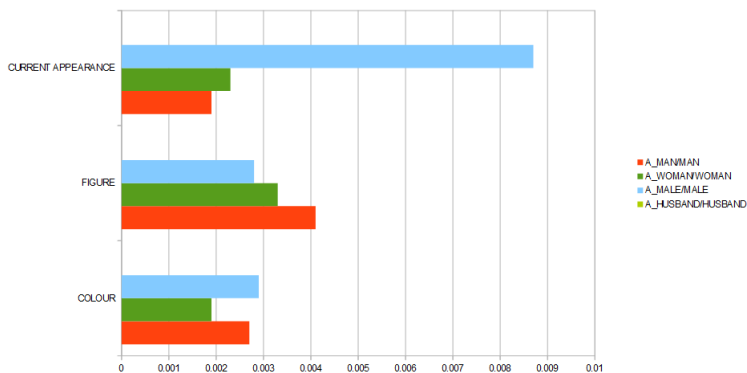


Figure 2. Physical appearance

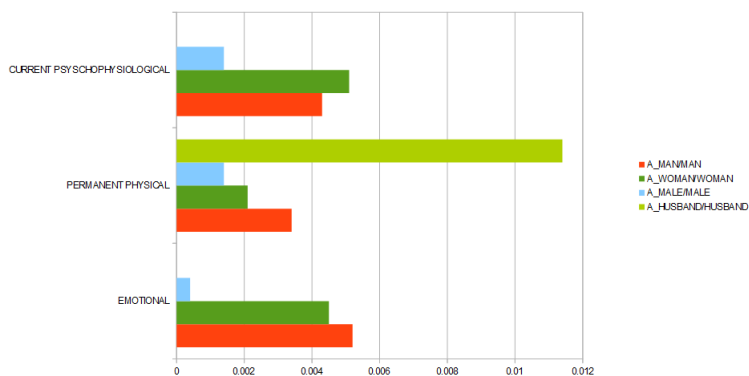


Figure 3. State

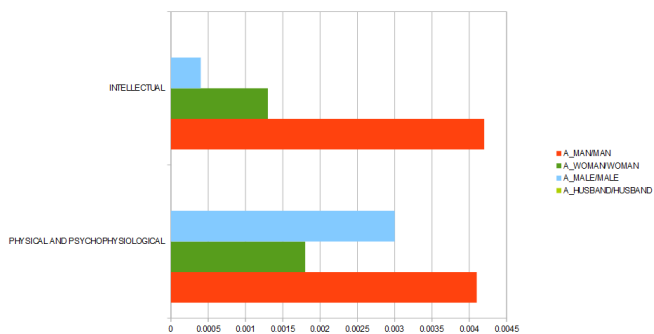


Figure 4. Predispositions

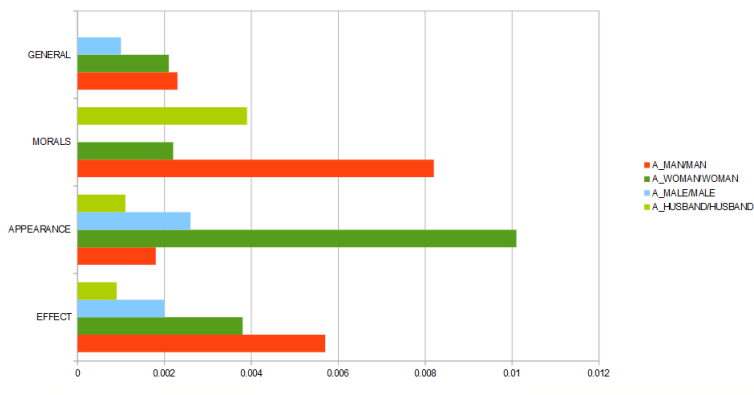


Figure 5. Subjective evaluation

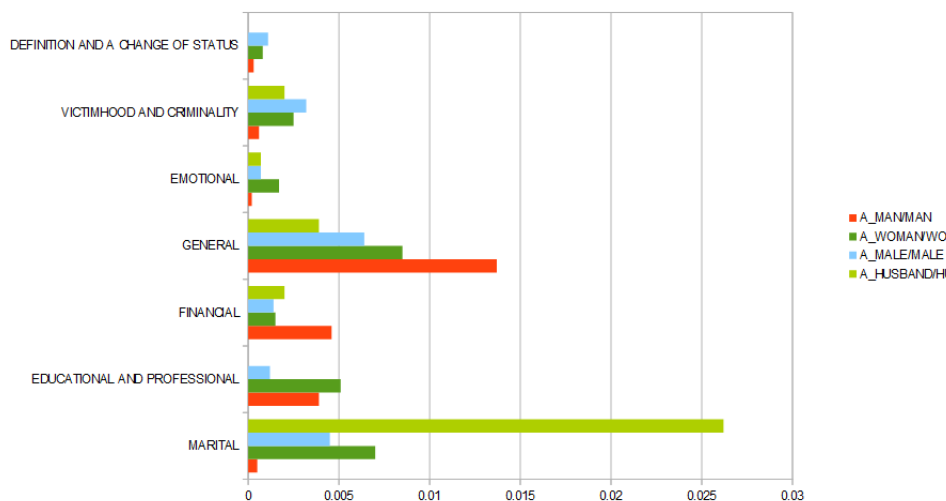


Figure 6. Social and personal status

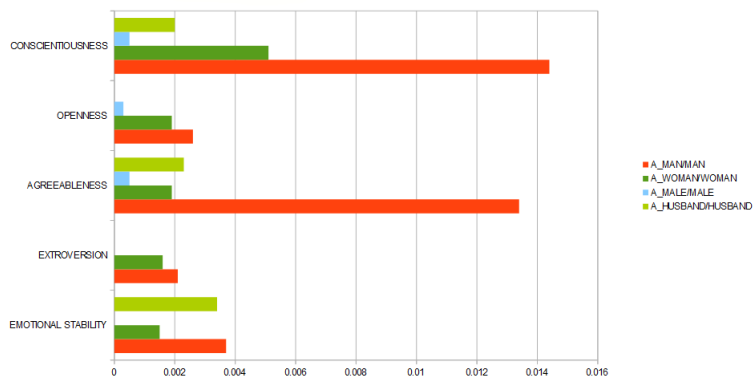


Figure 7. Personality model

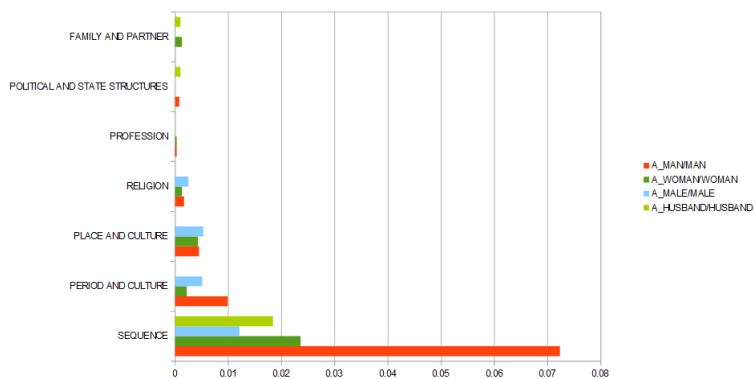


Figure 8. Belonging

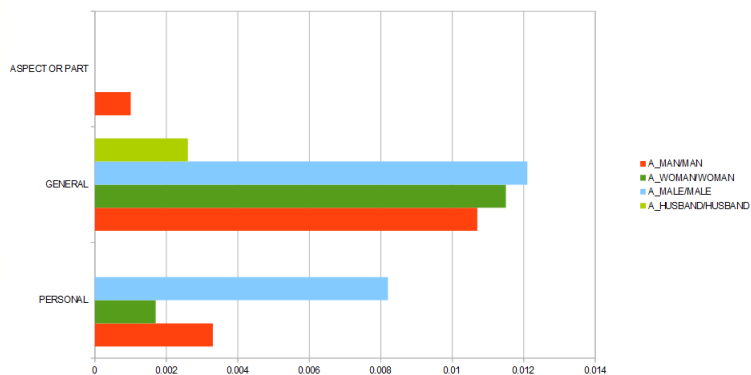


Figure 9. Relations

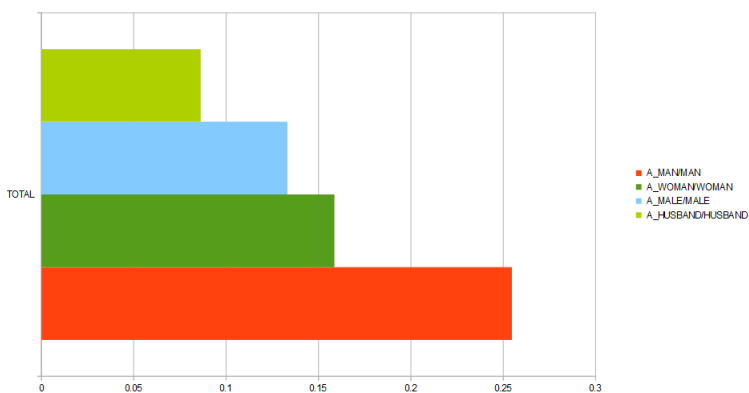


Figure 10. Overview of the overall distribution

Working on the Multimedia Document

“Al’ se nekad dobro jelo”¹

UDC 004.55:378.147]:02(497.11)

DOI 10.18485/infodhca.2017.17.2.5

ABSTRACT: Multimedia documents is a subject at the Department of Library and Information Science at the University of Belgrade, Faculty of Philology. Its aim is to senior students create a joint project that will be one kind of the final work. In this way, students have the opportunity to take advantage of the knowledge they gained during basic studies. The multimedia document for the academic year 2015/2016 was entitled “Al’ se nekad dobro jelo”. The idea was to gather and show diverse content related to traditional cuisine in Serbia and the neighboring countries.

KEYWORDS: multimedia document, Department of Library and Information Science, traditional cook, food.

PAPER SUBMITTED: 16 July 2017

PAPER ACCEPTED: 2 September 2017

Andela Stošić

balerina2993@gmail.com

Marina Milošević

marinamilosevic993@gmail.com

Sandra Spasić

sandraspasic@hotmail.com

Dragica Dragosav

dragica_dr@hotmail.com

Ana Đorđević

anadj93@hotmail.com

Ines Stanić

kurumizawa.ines@gmail.com

University of Belgrade

Faculty of Philology

1 Prologue - Tasteful Memories

Food is a missionary completing our lives, giving us the necessary energy and encouraging our creativity. It is like the Sun, Air, or Water, it is the Source of Life. Our senses are meant to be used in the purest possible way, so why should we refrain from the most diverse scents and tastes. All living beings are eating, growing and developing. Food carries the Breath of Life. It can also be a very powerful medicine, provided it is eaten fresh and raw so

¹ The title of this project is inspired by Đorđe Balašević’s song “Back Then Eating Was Good”.

that it does not lose its precious ingredients. Food connects us and helps to get closer and become kinder towards each other.

I believe that food prepared in the old days was rich and tasty, because with each movement we create a lot. Our grandmothers prepared corn bread or wheat bread on a daily basis. They repeated this process so many times, that they have led that to perfection. They baked pies and made buns and fritters. We dipped the fritters into sugar and we thought that was original. A friend told me, with a lot of excitement, how her grandmother used to prepare a fantastic meal with dumplings, called "trganjci", which she would be able to smell as soon as she approached her grandmother's house. I am certain that every one of us can recall a certain food that causes similar feelings to resurface.

The food used to be varied. For the tongues and palate there is nothing more pleasing than the rich taste of honey and jam. Grease spread on the bread is still on the menu today. Despite everything we have at our disposal, many still prefer these simple, old recipes.

Cows were a real fortune. They gave us invaluable milk, which was used not only for drinking but also as an addition to various meals. From milk we produced cheese and "kajmak". "Proja" with yoghurt and "cicvara" are an all-time favorite. The mixture of corn flour and boiling water becomes a really tasty meal when a lot of cheese and "kajmak" are added. I remember that we would not throw old bread; instead, we would let it dry and then pour hot water over it in order to make "popara". We would eat this type of food with "kajmak" and garlic.

Soups were made with various vegetables, magical combinations of which resulted in wonderful meals. Potatoes were prepared in countless ways. If you ask me, the potatoes wrapped in corn flour and baked with a little spice are perfect. However, beans cooked in earthen pot are the preferred dish for most people, and a season salad gives it a special flavour. Salads were a key part of every good meal. Previously, people did not buy oil. Vegetables were cooked on pork fat. A tree of sour apples in the courtyard was enough to produce the most natural and the healthiest vinegar of the highest quality. We used it for salads and various stupes. The king of all spices was parsley. The benefits of the herbs are numerous, but the aroma that spice herbs give to food is a perpetual bliss.

When I got the idea to put my memories in writing, I talked to my family and close acquaintances to see what they remember. Mostly they told me that, while in the past people did not have much, they were still happy and had more friends. It did not matter to them whether they were eating

a soup, rice or some meal with meat, they were healthy and full of energy. Maybe they were dancing in the rain. People always worried about one thing or the other. Some say that all this is natural and perhaps necessary for a person to see what is really important. However, life is more beautiful when we care less.

My mother told me that she often went fishing with her grandfather. Man has been hunting for survival since the dawn of time. The habit of using flesh as food was kept even when some animals had been tamed. Animals give us a lot, and their ultimate sacrifice is their flesh. That should cause gratitude and respect in us. And at some point, as we grow and evolve, maybe we will find that meat is not necessary for us to be strong and healthy and maybe we will establish a more honest and more intimate relationship with animals. Until then, we should provide them the most humane treatment possible. Some things are naturally transferred from the older generations to the younger ones, and it is up to us to accept what we love and want. It is as simple as that: we like different things.

The story of the sweets is an especially sweet one. Plums, apples, pears and strawberries, grown far from contamination, were a special treat. Fruits were dried, honey produced. Old women used to cook semolina with that basic sweet ingredient – sugar. Over time, they made a variety of cakes: gingerbread cookies, strudels, donuts, “urmašice”, “tulumbe”. “salnjake”, and so on. They almost always had the necessary ingredients for them: flour, fat, eggs and sugar. The variety of ingredients in the kitchen increased over time: cinnamon, chocolate, cocoa, carob, coconut, vanilla, may be common today, but in the past they were rare and thus more appreciated. Cakes were made with many chopped walnuts because the walnut trees were full of them. Every family cares for their loved ones in a unique way, but as far as my great-grandmother is concerned, she often made cakes for us and we considered her a sorceress because of that.

My friend recounted an event from her father's childhood to me: her father took some dry meat from the attic, and her good grandma scolded him for that because the food had to be used economically. A wise man once said that when you grow up modestly, you realize what you really need. We all have memories and only when they are not our burden, we can truly enjoy them and laugh wholeheartedly.

Today, a lot of different fruits and vegetables are used in the kitchen, but they are not produced as naturally. Cold pressed oils and seeds of different plants are used for preparing meals. Fruits that grow thousands of miles away, such as mango or coconut, are now within our reach. We eat while

we are working at a computer, walking and so on. My grandmother told me that in her time everyone used to have lunch together. Many grandmothers, like mine, did not go to school, and yet they are so mild and wise, that I can talk with them about everything. The truth is that everything is constantly changing, but the best things always remain unchanged, so it is interesting to see which of these things that inspired us once lives on today.

2 Preamble

The course "Multimedia documents" was introduced at the Department of Library and Information Science in the school year 2009/2010 and has been taught during the fourth year of undergraduate studies. The following themes that have been elaborated so far: Aleksandar Aca Popović (Тртовац, 2010), Kultne radio emisije (Cult Radio Programs) (Лазих и Поклопић, 2011), "Pop Ćira i Pop Spira" – Stevan Sremac (Priest Ćira and priest Spira) (Коврлија и др., 2012), Kako su nas zasmajavali - Radivoje Lola Đukić and Novak Novak (How Radivoje Lola Đukić and Novak Novak(ović) Made us Laugh) (Петровић и др., 2013), Žil Vern – Put oko sveta za osamdeset dana (Jules Verne – Around the World in 80 Days) (Перић и др., 2014), Igra rokenrol cela Jugoslavija (The whole Yugoslavia is Dancing rock'n'roll) (Обрадовић и др., 2016).

Professor Cvetana Krstev believed the most appropriate way for students to show what they had learned during their studies is for them to practically produce a multimedia document. This implies the application of knowledge related to the collection, publication and storage of materials in libraries, museums and archives, the ways of their processing, cataloguing and classification, and finally computer technologies, the Internet, information structures, programming and databases (Тртовац, 2010).

In the school year 2015/2016 Prof. Cvetana Krstev chose the topic "Al' se nekad dobro jelo" for the multimedia document for the entering class of 2012. The idea was to collect as much relevant material as possible for traditional cuisine in Serbia and the neighbouring countries. This project required the cooperation of students and people and institutions in their surroundings.

3 Cookbooks publishing in Serbia and the regional countries

Cookbooks publishing in Serbia and the regional countries To learn about publishing of cookbooks in Serbia and other countries in the region in different time periods, we searched the union bibliographic/catalogue database of our host system COBIB.SR and the regional databases COBISS.net

3.1 Searching the database COBIB.SR

The search was originally done by the keyword *kuvar** because through the short search by the keyword *kuvar** the results were obtained for both *kuvari* and *kuvarski recepti*. In this way, we retrieved 3,846 records.

Note: The subject heading *kuvar** does not enable retrieval of information in all relevant records because some of them do not contain data in block 6XX. However, searching in this way did not give us adequate results, because when we looked closely, we saw that many records did not refer to cooking recipes that we were looking for. For instance, among retrieved records were: "SQL kuvar" (SQL cook), "PHP kuvar" (PHP cook), "JQUERY kuvar" (JQUERY cook), "Kuvar: šta mladi umetnik može da nauči od advertajzera, mašine, budale?" (A Cook: what a young artist can learn from advertisers, machines, fools?) and so on.

For that reason, we used expert search through the freely chosen subject heading *kuvarski recepti* (cooking recipes).

The number of publications in a certain period of time is presented in Table 1, a total number of matches was 3,167.

Year	17*	18*	1900- 1918	1919- 1938	1939- 1945	1946- 1965	1966- 1985	1986- 2005	2006- 2016
Number of records	2	14	10	19	10	66	537	1362	1147

Table 1. Number of records in different time periods

It should be mentioned that the database also contains 53 records with the subject heading *kuharski recepti*. Those records were taken over from other databases without the adjustment of subject heading (*kuharski* to *kuvarski*).

We present some examples that we found interesting when searching the database COBIB.SR. These cookbooks were published during the First and the Second World War.

- РАТНИ кувар у којем су упуства за припремање јела већим делом готовљена без меса а за ове ратне прилике : ручна књига за наше домаћице. - Ујвидек : С. Ф. Огњановић, 1915 (Ујвидек : Задружна штампарија). - 48 стр. ; 16 цм. - (Народне књиге ; 1)
- ЗИМНИЦА. - Београд : ИПРОЗ, 1942. - 47 стр. ; 19 см. - (Мали кувар ; 3)
- Пекмез, мармелада, џем : написао Властимир Црнчевић. - Београд : ИПРОЗ, 1942 ([б. м. : б. и.]). - 39 стр. : илустр. ; 22 см. - (Пољопривредна библиотека ; 22)

In the 1950s a large number of cooking handbooks and textbooks containing recipes appeared, because various cooking schools were opened. Vegetarian cookbooks also appeared. The most interesting example is "Veliki narodni kuvar" (The Great National Cookbook) by *Spasenija Pata Marković* released in 1956. This cookbook, best known as "Patin kuvar" (Pata's cookbook) had as many as 29 editions, all published by *Narodna knjiga*. In the sixties, Pata's cookbook was also issued in the Braille.

In the period from 1966 to 1985, a large number of cookbooks presenting foreign cuisines (English, Hungarian, German, Japanese and Chinese) appeared. These are the years when many young people moved to the cities in search of work, usually living as bachelors; that is why a cookbook by Aleksandar Bubanović published in Zagreb 1985 was called: "Prva kuvarica za samice i samce : početnicima i naprednima" (The first cookbook for bachelors: beginners and advanced).

Cookbooks have also been following technological innovations. In 1984 in Ljubljana the following cookbook was published: "Kuhajmo v mikrovalni pečici" (Let's cook in a microwave oven).

The change in food preparation is also reflected in specific cookbooks that were published in the 1990s and in the beginning of the 2000s – cookbooks for macrobiotic cuisine, healthy food, meatless meals. Two interesting examples are "Frojdov kuvar" (Freud's cookbook) from 2005, translated from English, and "Kafa : 101 recept za ljubitelje kafe" (Coffee: 101 recipes for coffee lovers) by Ljiljana Andrejević, 1999.

The tendency to lead a healthier lifestyle can be observed in recent years and is reflected in cookbooks that promote use of raw and gluten-free groceries. One of the examples is a publication intitled "Presna hrana

za početnike : vodič za vitalnost i zdravlje putem prirodne presne veganske hrane” (Raw food for beginners: a guide for vitality and health through natural fresh vegan food) by Margareta Vesna, published in Belgrade in 2016.

3.2 Searching the database COBISS.net

We searched the database catalogues of the neighbouring countries Bosnia and Herzegovina, Macedonia, Slovenia and Montenegro and the query was a freely chosen subject heading *kuvarski recepti* (with equivalents in different languages of the regional countries) from 17th century to 2016.

Country		Macedonia	Bosnia and Hercegovina	Slovenia	Montenegro
Time period	17th and 18th century	0	0	28	0
	1900-1918	0	1	18	0
	1919-1938	1	0	26	1
	1939-1945	0	0	13	0
	1946-1965	2	8	105	6
	1966-1985	5	90	504	22
	1986-2005	14	322	4908	348
	2006-2016	20	110	2954	283
	Total	42	531	8556	660

Table 2. The number of records in the database COBISS.net obtained by expert search with the subject heading *kuvarski recepti* (and equivalents in languages of different countries) from 17th century to 2016.

Montenegro

By searching the catalogues of libraries in Montenegro we obtained 660 records and we can see the results in Table 2. Some of the interesting examples are:

- *Ukusi regiona : izabrana jela Jugoistočne Evrope i Balkana / Stevo Karapandža, Janez Bogataj ; [fotografije Janez Pukšić]. - Podgorica : Studio Moderna, 2011 (Novi Sad : Foto-OKO). - 115 str. : ilustr. ; 24 cm*
- *TRADICIONALNA jela iz Crne Gore / [priredile Anka Žugić, Vukica Knežević]. - Podgorica : Organizacija žena Crne Gore, 2007 (Cetinje : Obod). - 108 str. : ilustr. ; 20 cm*

- *Novi praznični kuvar / [pripremila Blanka Radošević-Marović]. - Podgorica : Pobjeda, [2006] ([s. l. : s. n.]). - 80 str. : ilustr. u bojama ; 17 x 24 cm*

Bosnia and Herzegovina

We used the Boolean expression *kuvarski* or *kuharski recepti* because records used one or the other or both subject headings and we obtained 531 records. Some of the examples are:

- *Bosanski kuhar : tradicionalno kulinastvo u Bosni i Hercegovini / Alija Lakišić ; [ilustracije Pavle Bugarin]. - Sarajevo : Svjetlost, 1979. - 462 str., pril. : ilustr. ; 25 cm*
- *HRANA za mozak : jelovnici za zdrav mozak i dugo pamćenje / Vida Demarin i suradnice ; [ilustracije Ivan Šarić]. - Sarajevo ; Zagreb ; Beograd : V.B.Z., 2010 (Ivančna Gorica : Impress). - 148 str. : ilustr. u bojama ; 20 cm. - (Biblioteka Stručno popularna)*
- *PRAVILNOM ishranom do zdravlja / priređivači Melida Hasanagić, Izet Mašić. - Sarajevo : "Avicena", 2011 (Fojnica : "Fojnica"). - 144 str. : fotogr. u bojama ; 24 cm. - (Biomedicinske publikacije ; knj. 36)*

Macedonia

The Macedonian database yielded 42 records through the search request *gotvarski recepti*. It has already been mentioned that the base also contained 19 records with the subject heading *kuvarski recepti* and five records with the subject heading *kuharski recepti*. These include:

- *СЛАТКИ за деца / [превод од српски Ирена Добреска]. - Скопје : Просветно дело, 2011 (Тетово : Напредок ; Крагујевац : Графостил). - 62 стр. : илустр. ; 24 см. - (Рецепти злато вредни)*
- *ТОРТИ, куглофи и тартови. - Скопје : Просветно дело, 2011 (Тетово : Напредок ; Крагујевац : Графостил). - 62 стр. : илустр. ; 24 см. - (Рецепти злато вредни)*
- *ПИТИ и солени торти. - Скопје : Просветно дело, 2011 (Тетово : Напредок ; Крагујевац : Графостил). - 62 стр. : илустр. ; 24 см. - (Рецепти злато вредни)*

Slovenia

In order to search the catalogue we used the Boolean expressions *kuvarski* or *kuharski recepti* because records used one or the other or both subject headings (su=kuvarski recepti yielded 245 hits, while su=kuharski recepti yielded 8,277 hits)

We highlight some interesting examples:

- 50 idej. Juhe in enolončnice iz morja / [prevedla Marjana Samide ; fotografije Andrea Fantauzzo, Dario Grimoldi]. Ljubljana : Delo revije : Pisanica, 2003 (Firenze : Bonechi). - 63 str. : ilustr ; 20 cm
- RECEPTI za zdravo srce. - Izd. za Crnu Goru. - Beograd : Mediasat East Europe ; Podgorica : Daily press - Nezavisni dnevnik Vijesti, 2006 ([Ljubljana : Mkt Print]). - 63 str. : barvne ilustr. ; 22 cm. - (Chef express ; 13)
- Čokolada / Andrej Goljat ; [fotografije Cveto Sonc]. - Ljubljana : Kmečki glas, 2006 (Ljubljana : Schwarz). - 117 str. : ilustr. ; 27 cm

By looking at results presented in tables 1 and 2 it can be seen that the publishing production with culinary themes has significantly increased after the Second World War, while the richest period is from 1986 to 2005. The exception is the Macedonia where most of the publications come from the last ten years.

Apart from books, a lot can be learned about food preparation from the media – television and the Internet. Therefore, in the next section we deal with culinary shows, blogs and web sites.

4 Cooking on TV and the Internet

Food is one aspect of culture. Everyone seems to have something to say about it. Today in a short period of time, we can talk about food, read about food, eat food, and watch how food is being prepared. Information about dishes, specialties and restaurants are interwoven on the Internet and in many printed publications, such as books, magazines, or newspapers. Cooking is no longer considered hard work, but rather a form of entertainment and a way of connecting. Food connects people and, at the most basic level, it can prompt a simple and friendly conversation. Whether it is a discussion of your current menu or about where you want to eat next week, food is a common ground. In addition, with a wide choice of food programs on television, Internet and in printed publications, it is easier than ever to find a recipe for a delicious meal.

Television shows about cooking are becoming complex, as recipes are not their only content. They are interwoven with educational and creative ideas and interesting workshops, and their audiences are, in turn, becoming more and more diverse, and everyone seems to be interested in food. Not a day goes by without at least one cooking show running on the TV. If you take Wednesdays for example, major domestic TV channels broadcast at least five

culinary shows. These include: recipes from Tito's kitchen (beyond the reach of many) on B92, reality show "Dođi na večeru" (Come to dinner) on the TV channel Prva, "Gastronomad" (Gastro-nomad) on RTS1. If we add to that shows on foreign channels that are broadcasted either in a language closely related to Serbian or subtitled, on Wednesdays the local audiences can watch at least twenty culinary chows. For instance, HRT and OBN broadcast shows "Slatko ludilo" (Sweet Folly) and "Stol za 4" (Table for 4), while channel TLC has at least five different shows per day (Dukić, 2013).

One of the first shows of this type was produced by the company "Advance" 15 years ago in cooperation with Vojislav Voki Kostić, a composer, a writer and a cook. Today, their "Gastronomad" is one of the most popular short television formats in the country. This does not come as a surprise because this trend is present all around the Globe, and TV cooks are gaining in popularity. TV culinary shows can be a way to expand cultural and social knowledge through, most obviously, culinary knowledge. The viewers are informed about different tastes, and learn how to recognize social differences through their taste of food. However, even with a most interesting scenario and great recipes culinary shows would not be as successful without a charismatic television presenter. Production houses in the West sometimes search for the right presenter for years. Sometimes they get lucky, as in the case of Jamie Oliver, who was accidentally spotted in the documentary "Christmas at River Café", where he worked as a cook. As soon as the BBC noticed his talent for public appearance, the famous show "The Naked Chef" was launched, and the book with the same title became the bestseller in the UK in the same year. In our country, one TV personality that is favorite among lovers of good food is Nenad Gladić, Alias Lepi Brka, who, besides hosting the aforementioned famous show "Gastronomad"², is also the editor-in-chief of the Kitchen TV. In addition to the traditional culinary shows Lepi Brka is also engaged in the "Best from Serbia" program on a cable channel for the food related events.³ The economic, touristic and cultural contents of these manifestations represents an extraordinary sample of the environment and promotes national creativity, as well as events in agriculture of a local and regional character. It is the complex cultural and socio-economic events that, among other things, cause greatest media coverage and public. There are numerous manifestations such as The Bacon Festival in Kačarevo, The

² Gastronomad, <http://www.gastronomad.rs/>, accessed March 30, 2017.

³ Najbolje iz Srbije, <http://kitchentv.rs/najbolje-iz-srbije/>, accessed March 30, 2017.

Cabbage Festival in Mrčajevci, The Sausage Festival in Turija, accompanied with the events dedicated to grape harvest in wine-growing regions and many other events.

Following the success of popular television channels like 24Kitchen, the Advance Media production house launched Kitchen TV, the first local and regional culinary channel to broadcast a large number of humorous culinary shows. One of these shows is called “Lonci i poklopci” (Pots and Lids)⁴, presented by Vjekoslav Kramer, the most famous chef from Bosnia and Herzegovina. This amusing culinary show is authentic and interesting just like the food prepared. Vjekoslav Kramer prepares simple meals in a completely new and specific manner, and devotes special attention to the names of the meals he cooks.

Another show broadcasted on Kitchen TV is called “Sa slatkim na ti” (With sweets on ‘thou’), where we are led through a fairy-tale world by Petar Žuljević, one of the leaders of the Crowne Plaza hotel team.⁵

In addition to classic cookbooks in printed form where the desired recipes can be found, webpages with recipes are also becoming numerous. Some of the popular websites, providing a wide range of sweet and salty dishes, are Coolinarka (<https://www.coolinarka.com/>), “Oklagija” (Rolling pin) (<http://www.oklagija.rs/>) and “Recepti za torte” (Recipes for cakes) (<http://receptizatorte.rs/>).

In recent years more and more web sites have been dedicated to the culinary domain, but culinary blogs have become even more popular. The best culinary blogs bring interesting recipes, explore cooking and analyze both foreign and domestic cuisine. Food blogs are very different from websites. They are made by individuals who are interested not only in cooking or trying out new dishes, but also in visiting the places that have different cuisine. Blogs are appealing to users who are eager to learn something about a recipe, food or a kitchen, as they can get detailed information on the blog. Blogs provide a variety of information about particular types of food, but the most attractive part of the blog are pictures. Feedback is also important because readers like to convey their experiences in the form of comments after they tried and tested a certain food recipe, and they also enjoy reading the experiences of others.

⁴ Lonci i poklopci, <http://kitchentv.rs/lonci-i-poklopci/>, accessed March 30, 2017.

⁵ Sa slatkim na ti, <http://kitchentv.rs/sa-slatkim-na-ti/>, accessed March 30, 2017.

"Zakuvavanje" (Boiling)⁶ is a gastro-blog where culinary fans can learn more about food, sometimes from a slightly different angle, because although women are traditionally viewed as housewives, men are here in general shown to be better chefs because they use everything they find and almost never follow recipes. Here, cuisine lovers can learn something more about well-known and lesser known recipes and cooking techniques. Yet another blog with an interesting topic is the "Brza hrana" blog (Fast Food)⁷. Different than all other blogs is the "Vitki gurman" blog (Slim Gourmand)⁸, hosted by Maja Petrović, where information about healthy nutrition can be found, as well as advices on how to introduce healthy nutrition in our ever faster and more stressful way of living.

On August the 1st, 2008, the woman blogger Monsoon started the first game for bloggers, which could be played by all interested people who would photograph and then upload their work on the Internet. After the second round, the game brought together 12 women bloggers, who named it "Ajme, koliko nas je" (Wow, how many of us are there). The game is played so that a game host for a certain month, through her blog choses a key ingredient, that is, the grocery that participants need to use. Some time during the month, the game participants have to make a dish with a given grocery and to publish it on their blog, with links to the founder of the game and to the host's post. A host publishes all works on her blog, selects the winners, and chooses a new host for the next month. She can also, if she wishes, send winners a gift of her choice.

The enthusiasm of this game led to the creation of a special culinary blog called "Kuar igrice" (Cooking games)⁹, where bloggers set up different games for those who are ready to cook meals, take photos of what they have done, and post a recipe of a produced and photographed meal.

It is obvious that food has always been a popular topic in everyday conversations, so it is not surprising that it was chosen to be the subject of our multimedia document. People enjoy talking about cooking and food, and this leads us to the idea that food, more than anything else, can be an incentive to conversation. Everyone needs to eat and should not be a special gourmand or expert to engage in a conversation about which restaurant has the best food or which cooking technique produces better results regarding

⁶ Zakuvavanje, <http://zakuvavanje.rs/>, accessed March 30, 2017

⁷ Brza hrana, <http://brzahrana.rs/>, accessed March 30, 2017.

⁸ Vitki gurman, <https://vitkigurman.com/blog-vitki-gurman/>, accessed March 30, 2017.

⁹ Kuvar igrice <http://kuvarigrice.blogspot.rs/>, accessed March 30, 2017.

to taste. Food plays special role in many celebrations and is important for marking special occasions for family and friends. Family recipes, favourite food shows and adventures in the kitchen - all this provides important facts about us and people around us.

5 A multimedia project - recording and photographing recipes, handwritten cookbooks, old kitchen items

Twenty eight students of the final year of the Library and Information Sciences graduate studies at the University of Belgrade were involved in the multimedia project "Al' se nekad dobro jelo". Their names can be found on the project's website: http://www.fil.bg.ac.rs/mmd_27/mmd_2016/studenti.html. In addition to the subject professor Prof. Cvetana Krstev, the project was coordinated by the assistant professor Miloš Utvić and the teaching associate Biljana Lazić.

The initial task was to find and photograph old kitchen items, appliances, cookbooks, etc. All students were asked to look up for them in their own households and consult their relatives, friends or neighbors.

For the same reason it was necessary to visit and explore the collections of the Museum of Science and Technology¹⁰ and the Ethnographic Museum¹¹ in Belgrade. The additional idea to find old restaurant menus in libraries had to be dropped since the libraries in Belgrade whose catalogues have been consulted do not keep menus in their collections.

Biljana Lazić distributed the tasks in agreement with the students: some were in charge of the cooperation with culture institutions, while the others processed video materials. One group of students dealt with the site design, the other was in charge of database creation. Some students searched for original music from regions where certain recipes originated.

5.1 Cooperation with individuals

All students cooperated with their family and neighbors who were assigned a task to prepare a traditional meal of their choice, and they made a great effort to record every detail of the preparation. Depending on the possibilities (technical equipment, readiness of the interlocutor to be recorded, etc.), some students made films, while others made photographs, from which

¹⁰ <http://www.muzejnt.rs/site/home/>

¹¹ <http://etnografskimuzej.rs/>

presentations were later compiled in the form of short video clips. In this way, students had to collect at least two dishes and one more that they would prepare on their own.

Apart from the students' household members, even relatives contributed to the project, along with neighbors and all those who were enthusiastic about the old ways. In conversation with them, nostalgia was in the air. As they say, they did not have much, but they shared a lot. The children looked forward to a simple sugar cube, and each treat was a real rarity. Many of the interviewees pointed out that once upon a time, one appreciated a lot what little one had, while today, when we have more things, we fail to see the value of small things. People also used to care what was eaten and how it was being prepared. Old recipes represent the knowledge gathered over the past decades, sometimes even centuries. The notes we gathered meet two human needs in a wonderful combination – a need for knowledge, and a need for food.

The following people cooked for our project: Aida Bazdar, Anđelka and Marijana Vasić, Božica Čairović, Branko Lukić, Vasil Vasilev, Verica Dragić, Diana Kesić, Dragana Milošević, Dušanka and Željko Ivanišević, Zagorka Cvetanović, Zoran and Ljiljana Skorupan, Zorica Dragosav, Zorica and Milica Prvulović, Radica Simonović, Ružica Đurašević, Simona Čirović, Slavica Purić, Slavica Stošić, Slađana Spasić, Slobodanka Stanić, Snežana Milošević, Snežana Stevanović. Their names and short biographies are listed beside their recipes on the Multimedia Document website.

In addition to two printed cookbooks (*The old and the new kitchen* from 1877, *The Great National Cookbook* from 1956), our multimedia document contains old handwritten cookbooks with recipes collected in the mid-20th century. These handwritten cookbooks resembled scratch pads. Some pages being worn, students avoided touching them so they would not fall apart, and it was difficult to interpret old handwriting. All notes given to us by the housewives were scanned and translated into the digital form, and can be found and browsed on the web pages of the project. The authors of these notes are Tatjana Kostić, Stamenka Gašić, Žanja Petković, Metka Đorđević, Nada Milosavljević, Ankica Dukić and Jelena Zebić. Thanks to them and their desire to help, we did not only obtain recipes, but also learned about many customs and habits from the past that were part of the secret skills of our cooks.

Since the collection and preservation of objects from the past is a fundamental necessity, the students came across people who, in their private collections, held various artifacts, objects from the old times, the rarities

that housewives once used in their kitchen. Housewives were eager to show items they inherited from their grandmothers, mothers or mothers-in-law. Tamara Dukić, Tatjana Kostić, Nevena Prvulović, Marina Milošević, Irena Ajruloska, Maja Tadić, Milica Radić and Aleksandra Trtovac photographed and presented old dishcloths, mortars and pestles, schnapps jugs, coffee and pepper mills, fruit and vegetable presses, machines used to crush cherries, old plates and bowls, and many other items unknown or not in use today. These objects are predecessors of items in use today. This project allowed us to remember them. Interviewees who shared their recipes and experiences with us, gave us the permission to display various old objects from their attics, cellars, sacks and other places that shelter the spirit of the past days.

5.2 Cooperation with museums

In order to present the objects from both museums in the best way, students were grouped in two groups. Students Sandra Spasić and Dajana Lazarević went to the Ethnographic Museum, while Anđela Stošić, Slađana Ivanišević and Ana Đorđević visited the Museum of Science and Technology. Students were guided through the permanent exhibition of the Museum of Science and Technology. This experience enabled them to discover old stoves, refrigerators, ice boxes, mortars and pestles, copper molds, milk bottles from Serbia from the sixties, express pots from the first half of the twentieth century, napkin holders, dough molds, measuring cylinders, glasses, mugs, mixers, mills, teapots and many other items that were once used in everyday life. In the Ethnographic Museum students found folk costumes, travel baskets, accessories for curd processing, accessories for fishing and fish processing, earthenware, old schnapps pots, a wine cellar, kettles and many other items pertinent to the preparation of food and beverages. Our project also reveals the interior of the 19th century town, country house and kitchen, as well as how the festive table of a Serbian family looked like.

Miloš Jurišić, the photographer of the Museum of Science and Technology, a collector of old photographs and postcards, was of great help. Some of his photographs are included in our project. Old Serbian cafés and restaurants were different from the contemporary ones in their interior, prices and types of meals. Although we did not find old menus in libraries, Mr. Jurišić allowed us to include in our document some old menus he managed to collect in recent years:

- Menu for the wedding of Mrs. and Mrs. Leposava and Nikola N. Krstić from the Kolarac brewery;

- Menu for the celebration marking the opening of the University, on October 2, 1905;
- A menu from 1921;
- A lunch menu in honor of graduate cadets in the Palace of the King Alexander I, 1940
- Menu for the dinner of friends gathered after 25 years, on December 3, 1939, at the Hotel Srpski kralj, Belgrade.

6 Designing and creating a website as a final part of the project

In order to create a database and a web site containing photographs, audio and video materials, texts, scanned cookery books (whether handwritten or printed) that were gathered in the scope of the project, students had to apply the knowledge of the information sciences acquired during their undergraduate academic studies at the Department of Library and Information Science, University of Belgrade.

This multimedia project is important because it combines the old and the new, the Industrial and the Information age as well as the Digital age. Old Serbian dishes, customs and habits, accompanied by the traditional music, received their digital form and were presented on the site designed and maintained by the students. Visitors have the opportunity to access the digital content about the traditional Serbian cuisine at http://www.fil.bg.ac.rs/mmd_27/mmd_2016/pocetna.php. Included are various multimedia documents such as photos, video and audio recordings, scans of old cookery books etc.

Students were organized in two groups based on their interests, knowledge and skills. One team was responsible for the creation, structuring and design of the web pages, while the other team was in charge of the database and PHP¹² programming that would offer better recipe search. The teams worked together and exchanged ideas that led to planning and implementation step by step. The development of the site architecture and of the database was a very complex task where every team member had his/her responsibility. The students worked on the project independently, but the project mentors were there to help, to offer suggestions and advice, support them and make their work easier.

¹² PHP (Hypertext Preprocessor) is a server-side scripting language designed primarily for web development.

6.1 Video processing

The most extensive part of the Multimedia Document are culinary recipes, 98 in total, recorded in mp4 format.

As already mentioned in the previous section, some students recorded the entire cooking process, while others just took photos. Students who could not record their cooking associates had to produce video clips from their photographs. The VideoPad Video Editor¹³ – a free program for producing videos for beginner, turned out to be the most suitable one. In addition to this program, students used the SuperDVD Video Editor¹⁴, another video-making program. Some students also used the Windows Live Movie Maker¹⁵ program to produce video recordings and create media presentations in a simple way. As it was universally agreed that the project should be as visually attractive as possible, the students gave their best and all photographs were converted to videos. In the team dealing with the creation of the database and PHP programming were Anđela Stošić, Dragica Dragosav, Marina Milošević and Ana Đorđević. They were in constant cooperation with the Assistant professor Miloš Utvić and the web design team. Students decided to include all video recipes in the database and classified them according to three criteria:

- Origin – Vojvodina, Belgrade, Šumadija and West Serbia, South and Eastern Serbia and Kosovo and Metohija, Bosnia and Herzegovina, Croatia, and Slovenia¹⁶;
- Ingredients – weather a particular meal is suitable for fasting
- Types of dishes – dishes were grouped along the sections of Pata's cookbook (Марковић, 1956) – appetizers, salted pies and dough, soups and broths, cheeses, roasts, salads, cakes, desserts.

The main table in the database is recepti (recipes) (Figure 1) and it contains the following columns: nazivJela (name of a dish), jeloID (dish ID), tjvlID, regionID, tjlpD. The primary table key is jeloID values of which are three-digit numbers unique for each recipe and used for naming recipes in a folder. The tjvlID is a foreign key to the table tipovijelaverski (types of dishes),

¹³ <http://www.nchsoftware.com/videopad/index.html>

¹⁴ <https://superdvd-video-editor.en.softonic.com/>

¹⁵ http://download.cnet.com/Windows-Live-Movie-Maker/3000-13631_4-10965753.html

¹⁶ Most recipes were collected in the territory of Serbia, but there are also few recipes from neighboring countries.

indicating whether a dish is suitable for fasting. The `tjpID` is a foreign key to the table `tipovijelapata` (types of dishes). The `regionID` is a foreign key to the table `regioni` (regions).

jeloID	1	nazivJela	tjpID	tjvID	regionID
100		Pasulj	23	3	13
101		Slane pogačice	21	3	13
102		Štrudla	27	3	13
103		Koh	27	3	11
104		Domaća baklava	27	3	12
105		Domaća pita	21	3	12
106		Štrudla	27	3	12
107		Pile na podvarku	24	3	12
108		Prokule i sarma	23	3	12
109		Serbo kocke	27	3	12
110		Domaća testenina	22	3	12
111		Domaći burek sa mesom	21	3	12

Figure 1. Part of table `recepti`

All three divisions are used on the website as search parameters. When a site visitor chooses a type of food and/or a recipe origin, PHP script consults the database and formulates the SQL query that yields the desired result, a web page is generated showing the names of the dishes featuring required characteristics, the video clips showing the preparation, as well as information about the cook. The homepage offers the possibility to search recipes by name. It is necessary to enter at least three letters of a name of a particular dish and if there are matches with recipes from the database, the web page is generated (Figure 2).

The team working on web-design was led by Ines Stanić. Students had to express their ideas by using the markup language HTML, the CSS language for describing visual style of HTML documents, the JavaScript programming language. All of this was done in cooperation with the PHP team. In this project, the web-design involved the design of all pages and layout of their content, as well as encoding. The code was written in Notepad++¹⁷ and

¹⁷ <https://notepad-plus-plus.org/download/v7.4.2.html>.



Figure 2. An example of the search result

photographs were edited in SketchBook¹⁸. The website is simple and easy to use, with clearly separated content, links and search fields. Students found very helpful to investigate free access codes and websites that are intended for studying web design and programming.¹⁹

The visual style of webpages and text was based on color psychology. Green and red were chosen to be the most dominant. Green color symbolizes the nature and health, and red, as diametrically opposed, stands for activity, joy of life and appetite (Trstenjak, 1987). Also, it is important to stress that the photographs used for the website were originally produced by students during the project, while the music used for the audio effects belong to various authors and are used in accordance with the copyright laws (see the Note at the end of the article).

7 The project for all generations

Traditional cooking, as a topic, gathered all generations in the students' environment, the older people were happy to recall their childhood days,

¹⁸ <https://www.sketchbook.com/?locale=en>.

¹⁹ The most useful were: <https://www.w3schools.com/> and <https://stackoverflow.com/>.

while the younger people were grateful for listening about unforgotten experiences from the past. Almost forgotten tastes, scents and customs were awakened and all have enjoyed it.

Although this is an unusual exam, it still demands a lot of effort and time, and extraordinary organization skills from everyone involved in the project. The multimedia document is a final project that includes students' knowledge acquired during the four-year studies. Courses such as Library Public Relations and Library Management have made communication with the institutions easier, while the course Organization of knowledge in Libraries was useful for classification of collected materials. The width of knowledge in IT subjects helped us to accept and master new software solutions. Motivation, initiative, as well as success in team work is something that this project has enriched us with.

8 Acknowledgements

At the very end, we would like to emphasize that this project would not be successful without our mentors, above all, and then numerous individuals and institutions that participated in the project. In this way, we would like to thank professor Cvetana Krstev, professor Miloš Utvić and their associate Biljana Lazić, for their effort, patience and desire that students improve their knowledge though this creative and comprehensive task.

We also thank all the associates whose names have already been mentioned in the paper, who participated in the preparation of meals or provided the items included in this multimedia document.

We owe gratitude to Ethnographic Museum and Museum of Science and Technology in Belgrade for professional guidance and interpretation of their permanent exhibitions.

Note

In order for the viewer to fully understand the atmosphere of each geographical area from which the recipes were collected, the following eight songs are included in the web presentation of the project:

- “Ja sam rođen tamo na salašu” (music: Danilo Danilović, text: Miomir Petrović),
- “Crveni šal”, sang by Dubravka Nešović, 1973.

- “Noć je tiha”, sang by Predrag Živković Tozovac, author of the song Živadin Jovanović,
- “Niška banja, topla voda”, sang by Olivera Katarina,
- “Ajde Jano”, sang by Mara Đorđević,
- “Bosa Mara Bosnu pregazila”, sang by Nada Mamula,
- Instrumental of the song “Ribar plete mrižu svoju”,
- “Slovenija”, sang by Ensemble Avsenik.

Beside this compositions, the song “Al’ se nekad dobro jelo” by Đorđe Balašević is included too, because, as we said at beginning, its title inspired the name of this project.

References

- Dukić, Marija. “Zašto su u eri ‘brze hrane’ popularne emisije o kuvanju”, *Biznis i finansije*, 2013. Accessed July 20, 2017, <http://bif.rs/2013/04/zasto-su-u-eri-brze-hrane-popularne-emisije-o-kuvanju/>
- Radunović, Luka. “Psihologija boja u marketingu”, *Digitalizuj me*, 2015. Accessed July 30, 2017, <http://digitalizuj.me/2015/06/psihologija-boja-u-marketingu/>
- Trstenjak, Anto. *Čovek i boje*, Beograd: Nolit, 1987.
- Коврлија, Дарија, Валентина Тасић и Сузана Топаловић. “Мултимедијални документ: Поп Тира и поп Спира - уједињено знање на заједничком студентском пројекту”. *Инфотека* Vol. 13, no. 2 (2012): 86–89. Accessed August 1, 2017, http://infoteka.bg.ac.rs/pdf/Srp/2012-2/INFOTHECA_XIII_2_December2012_86-89.pdf
- Лазих, Биљана и Јелица Поклопић. “Мултимедијални документ: Култне радио емисије - осврт на стање у архивима радио станица у Србији”. *Инфотека* Vol. 12, no. 2 (2011): 53–57. Accessed August 1, 2017, http://infoteka.bg.ac.rs/PDF/Srp/2011-2/INFOTHECA_XII_2_Decembar_52-57.pdf
- Марковић, Спасенија Пата. *Велики народни кувар*. Београд: Народна књига, 1956
- Обрадовић, Милена, Александра Арсенијевић и Михаило Шкорић. “Израда мултимедијалног документа ‘YU рок сцена’”. *Инфотека* Vol. 16, no. 1 (2016): 113–129. Accessed August 1, 2017, http://infoteka.bg.ac.rs/pdf/Srp/2016/infoteka-2016-16-1_2-6.pdf
- Перић, Катарина, Кристина Гогић и Ана Николић. “Израда мултимедијалног документа ‘Пут око света за 80 дана’”.

Инфотека Vol. 15, no. 2 (2014): 56–62. Accessed August 1, 2017, <http://infoteka.bg.ac.rs/index.php/sr-yu/arhiva/2014/2/infoteka-15-2-2014-56-62>

Петровић, Невена, Александар Николић и Агрон Бакиу. "Приказ мултимедијалног пројекта 'Како су нас засмејавали: Радивоје Лола Ђукић и Новак Новак(овић)'" *Инфотека* Vol. 14, no. 2 (2013): 49–61. Accessed August 1, 2017, http://infoteka.bg.ac.rs/pdf/Srp/2013-2/INFOTHECA_XIV_2_2013_49-61.pdf

Тртовац, Александра. "Мултимедијални документ као предмет-пројекат на Катедри за библиотекарство и информатику Филолошког факултета у Београду". *Инфотека* Vol. 11, no. 2 (2010): 45–54. Accessed 21 June 2017, http://infoteka.bg.ac.rs/pdf/Srp/2010-2/INFOTHECA_XI_2_December2010_45-54.pdf

Conference Proceedings “Legal Aspects of Digitisation of Cultural Heritage”

Dragana Stolić
stolic@unilib.bg.ac.rs

*University library
“Svetozar Marković”*

PAPER SUBMITTED: 5 November 2017

PAPER ACCEPTED: 4 December 2017

Proceedings of scientific conference held on 19. October 2017 at Serbian Academy of Science and Art “Legal aspects of digitisation of cultural heritage”, edited by Dragan Prlja,¹ contains fourteen original scientific papers which deal with the issues of “normative regulation of jurisdiction for digitisation of cultural heritage in Serbia, countries in the region and the European Union” from different aspects, as well as with many other questions concerning legal standards and intellectual property protection during the process of digitisation. In the foreword of the Proceedings it is pointed out that new possibilities of digitisation related to availability and protection of material brought new technical and organizational challenges too, but no less important legal ones. However, legal aspects have been neglected so far, although this activity has existed for twenty years. The aim of the Proceedings is to compensate the lack of literature, and to unite knowledge from this area in which the research will continue and in which some questions are going to get real answers.

Diversity of presented subjects and multiple approaches enabled the fulfillment of designated tasks and Proceedings provides comprehensive overview of documents which could be relevant for this issue, on the one side, and it discusses the question of protection and security of digitized material, from legal to technical in the strict sense. The Proceedings offers oversight of actual condition in national and international legal domain, but also indicates the aspects that should be improved and developed.

¹ Prlja, Dragan, ed., *Legal aspects of digitisation of cultural heritage*. Belgrade: Institute for Comparative Law, Serbian Academy of Science and Art, Audiovisual archive and Center for Digitization of SASA, 2017.

1 Documents

In the first article of Marko Reljanović and Dragan Prlja (“Digitisation of cultural heritage in the Republic of Serbia – normative aspects”) emphasizes the necessity of very well regulated normative framework concerning laws, but also the other legal documents. Although there are many law solutions in this area,² Law on archive documents and archive activity is yet to be enacted. The National strategy of development of information society in the Republic of Serbia till year 2020 treats the development of digitized content as one of the priorities. Changes of Law on copyright and similar rights, which are planned for the first quarter of the 2018 are very important and their goal is the harmonization of EU Directive on some permitted uses of orphan works. Then it will be possible, by following certain procedure, to digitize works whose copyright holders are unknown or unavailable.³

The overview of international standards of digitization, presented in paper of Ana Batričević, covers standards of UN, but also those of European Council and European Union.⁴ The activities defined within the goals of National center for digitization – such as the need for legal regulation of access to digitized content, followed by protection of intellectual property, right on commercial distribution of publicly available commercial content, mandatory deposit of digitized versions of new editions and contents born digital, rights for automatic harvesting of digitized content – are also emphasized.

² The most important laws are: Law on culture, Law on cultural works, Law on library and information profession, Law on cinematography, Law on electronic document, and above all Law on copyright and related rights. For this area it could also be relevant: Law on optical discs, Law on electronic signature, Law on free access to information of public interest, Law on electronic communications, Law on information security, etc.

³ “Orphan works”, or works of unavailable copyright holder, include works which are still under copyright, but whose copyright holders are unknown or for some other reason unavailable, so it is not possible to get the permission for making copies of the work. Many works from the first part of 20th century, which make part of cultural heritage, belong to this group.

⁴ Standards of United Nations include primary Charter of Unesco on preserving of digital heritage from 2003; the most important Standard of European Council is Convention on the value of Cultural Heritage for the Society from 2005, documents belong to EU standards are: Lund principles, Charter of Parma and Conclusion of European Council on the role of Europeana for the digital access, visibility and use of European cultural heritage.

List of national and international legislations are completed with presentation of legal aspects of digitisation in the region, mainly in Montenegro (Dražen Cerović), but also in Bosnia and Herzegovina (Iza Razija Mešević). In the paper of Jelena Kostić and Miloš Stanić the comparative law analysis of competencies of state authorities is given, during cultural heritage's digitization procedure in the Western Balkans countries. Short report of competences in each EU country is brought in the paper of Gordana Gamsi, Vanja Korać and Sanja Prlja ("European Union and digitisation of cultural heritage") in which the relevance of documents such as recommendations is heavily emphasized. Among them the Recommendation 2011/711/EU is particularly important since it defines "the rules of digitisation of cultural heritage in the line with the European Union Strategy by 2020", the implementation of which is monitored through periodical reports. The importance of Europeana portal, "large digital cultural platform" is underlined, through which in the years to come EU will endeavor "to provide progress in digitisation of cultural heritage", long term preservation, but also improvement of content quality and availability.

In the paper of Andrej Diligenski ("Court cases related to the digitisation of cultural heritage and its availability in cyberspace") the development of Europeana project is presented, thanks to which around 55 million works of cultural heritage are available today. Nevertheless, main part A. Diligenski's paper is aimed at court cases which often become legal frameworks for certain issues, since "they are live law sources and they improve shortage of law regulations". This is illustrated with three particular cases or verdicts brought in European courts: that books that can no longer be found in bookstores can be digitized and used only if their authors were aware of digitisation and had the possibility to cancel it, that the lending of digitized and printed copy in the public institutions is equaled and that libraries can under certain conditions make books available in electronic reading rooms without authors' permission.

Finally, in the paper of Jelena Matijašević Obradović and Joko Dragoljović documents like Digital agenda for Serbia are analyzed. Agenda, as initiative of Digital Agenda of Europe for the year 2020, is defined by the Strategy of development of Information society by 2020 and by Strategy of development of e-communication in Serbia from 2010 to 2020.

2 Protection and archiving of documents

Security aspect of digitized material is treated in the paper of Dragan Jovašević (“Security protection of the digital databases of cultural heritage of the Republic of Serbia”). In the paper it is emphasized that with the changes of Criminal law of the Republic of Serbia from 2003, based on international standards, especially Convention on cybernetic (cyber) criminal (2001), in the system of criminal law for the first time are brought “several cyber (computer) criminal acts”. This is incorporated in later version of Criminal law of the Republic of Serbia (2005), in which it is subsumed under section of Criminal acts against the security of computer data.

Protection of artistic works, as a part of cultural heritage which is also the subject of digitisation, can contain more technical aspects, which is discussed in the paper of Katica Tomić, focused on the liability in the process of authentication of art works and the implementation of new technologies (block chain), and in the paper of Nikola Paunović on the importance of databases of stolen cultural goods, for the prevention of illegal trade on internet. In this paper, the importance of Convention on criminal acts related to cultural works, which is adopted by the Ministry Council in 19th may 2017, is analyzed in particularly.

In the context of digitisation term of protection also indicated the issue of copyright protection, but this topic brings new challenges and traps, as it is discussed in the paper of Miodrag Savović and Danilo Rončević (“Digitisation between the protection of copyright and the right to use cultural heritage”). The difficulties with which institutions can face during the process of digitisation are indicated, which is related mainly to harmonization of authors’ rights on the one side, and public need of usage the cultural works by the wide range of people, on the other. Very often, institutions estimate the risk of digitisation and digitize even the material for which they do not have a permission. The conclusion brought in the paper is that the solution lies in constant changes and adjustments of legal regulations, in creating the database of authors which could facilitate finding of copyright holders, and in forming the competent councils who could provide necessary help.

The other, maybe ultimate challenge for each institution which starts the process of digitisation – “orphan works”, could be solved through so called model of collective rights management, as it is explained in the paper of Nataša Mrvić Petrović and Vladimir Čolović. This is a model to “bypass limitations of copyright”, named “ECL – extended collective license”, but it is applicable only on the territory of one country. Organizations for col-

lective management of copyright can become representatives of copyright holders, and they can freely negotiate with cultural institution for approval of digitisation. In different versions, this model is already implemented in Scandinavian countries, and in France and Great Britain.

Although in the majority of papers the accent is put on the conversion of analogue documents into the digital, and on application of legal solution in that part of process, in the paper on legal aspects of Digital repository of the University of Belgrade, Dragana Stolić and Tatjana Brzulović Stanisavljević, the attention is called on repository as means of preserving digitized material which makes the infrastructure for so called “green open access” and enables the use of digitized content. However, a few noticed problems are pointed out, such as: misunderstanding the way of law regulations in the frame of repository (Creative Commons licenses), doubts about ownership over deposited objects and, finally lack of accordance between different laws, mainly Law on copyright and related rights and Law on high education.

In spite of diversity of presented papers of the Proceedings, the dominant subjects in the context of discussion on law regulations related to digitisation of cultural heritage can be singled out. It is, above all, question of implementation of Directive EU on certain permitted use of “orphan works”, which will unblock digitisation process in one part and enable digitisation of those works which are still under the copyright. In many papers the importance of different ways of education is underlined, as well as the need for forming the instances for providing professional help. Legislation which regulates process of digitisation, directly or indirectly, is very wide in this moment, on national and international level, but these frameworks need to be constantly monitored, adjusted and completed if it is necessary, so that the requests posed by this long term and important process put in front of institutions can be fulfilled.

83rd - IFLA World Library and Information Congress

Jelena Andonovski
andonovski@unilib.bg.ac.rs

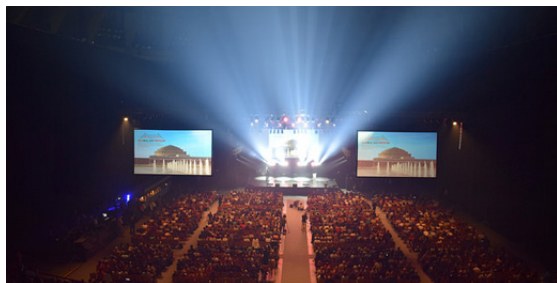
Nataša Dakić
dakic@unilib.bg.ac.rs

*University Library
"Svetozar Marković"*

PAPER SUBMITTED: 24 October 2017

PAPER ACCEPTED: 7 November 2017

The 83rd World Library and Information Congress (WLIC) called Libraries, Solidarity, Society was once again successfully organized by The International Federation of Library Associations and Institutions (IFLA). In the period from 19th to 25th August 2017, Wrocław, the historic capital of Silesia and the fourth largest city in Poland, was home for more than 500 speakers and 3100 delegates from 122 countries all over the world.¹



Serbian Library Association enabled free minibus transfer to Wrocław for 19 members of the Society, colleagues from all library types. Trip to Wrocław was realized due to the project "Librarians on the road: 1000 kilometers to 1000 ideas" co-financed by the Association's funds and project approved by the Ministry of Culture and Information. Thanks to this project, Serbian librarian's delegation with 30 registered participants was the largest Serbian delegation to IFLA congresses ever. Due to persistence and creativity, the

¹ <https://www.ifla.org/node/11814>

participants from Serbia received well-deserved attention, approval and tribute from the international library community, including now already former, IFLA President Donna Scheeder.



The colleagues who travelled to Wrocław thanks to the Serbian Library Association had the opportunity to visit different libraries in Slovakia, Czech Republic and Poland: University Library in Bratislava – the oldest and the largest library in Slovakia; Moravian Library in Brno – regional library of the South Moravian Region and the second largest library in the Czech Republic; Public library in Opole; as well as Silesian Library in Katowice – one of the most modern libraries in Poland and the scientific and regional public library.

After spectacular Opening Session of the 83rd Congress, with the warmest of hugs, Donna Scheeder handed over her presidency to Glòria Pérez-Salmerón. A new Governing Board was presented with a main task to steer IFLA in these exciting times of transparency, openness and fresh ideas.

Online voting platform “IFLA Global Vision” was launched with the aim to propel inclusive global discussion to unite the library field. Also, during the Congress, the real-time voting was organized where 360 delegates voted simultaneously. This voting platform had been opened until October 16th. Since the launch, the IFLA Global Vision has received more than 22 000 votes. Library experts from 213 countries gave their opinion about the place, role, challenges and activity of libraries in the contemporary world in the process of globalization. The project motto is “Together we create the future”

and we expect the final report with the guidelines for the development of the librarian profession in the future.

The congress itself was organized in 247 sessions which covered all aspects of our profession and represents an invaluable experience for all participants. It also represents a place to meet colleagues from all over the world, provides professional development and opens door for further collaboration between libraries. Many librarians from Serbia participated as volunteers, authors of papers and posters, members of sections and working groups of IFLA, and among them was one IFLA World Library and Information Congress Participation Grant holder.

The University Library "Svetozar Marković" was presented to the international library community very successfully through two posters and two papers:

1. Nikola Smolenski, Aleksandra Trtovac, Nataša Dakić, Jelena Andonovski, Adam Sofronijević: Adding value for society by efficiently presenting cultural heritage with minimum additional costs: In-house development of digital tools for searchable collections at University Library Belgrade (poster)
2. Marija Bulatović, Vera Petrović: University Library "Svetozar Markovic" in Belgrade and its digital funds (poster)
3. Tatjana Brzulović Stanisavljević, Dragana Stolić: Accredited Seminars as an Example of Good Educational Practice: Program 'Copyright in Library and Information Science' of the University Library 'Svetozar Marković' in Belgrade (paper)
4. Milena Kostić, Vesna Vuksan: Volunteer and Tech Power Bringing New Ideas to Life at the University Library, Belgrade (paper)

Participants from University Library presented the first fully searchable corpus of digitized historic newspaper texts with almost 600,000 pages in Serbian Cyrillic from the period between the beginning of the 19th century to the middle 20th century, as well as activities related to continuous professional development of librarians through the accredited courses, workshops and seminars.

The 84th IFLA World Library and Information Congress, called Transform Libraries, Transform Societies, will be held from 24th to 30th August 2018 in Kuala Lumpur, the capital of Malaysia.² The main feature and success of the Congress lies in the exchange of experiences, knowledge and skills

² <https://www.ifla.org/node/11754>



directed towards improvement of the profession as a whole. Therefore, we hope that librarians from Serbia will have the opportunity to participate.

Author Guidelines

All *Infotheca* articles are published both in English and Serbian in the same issue. Authors should submit their articles in one of the languages; only after the notification of acceptance the translated article is expected (for Serbian authors; for all other authors translation from English to Serbian is provided by the journal). Except the printed edition, all articles are also published in the online edition in open access.

PAPER CATEGORIZATION

For documents accepted for publishing which are subject to review, the following categorization in the Journal applies:

1. Scientific papers:
 - Original scientific paper (containing previously unpublished results of authors' own research acquired using a scientific method);
 - Review paper (containing original, detailed and critical review of a research problem or a field in which authors' contribution can be demonstrated by self citation);
 - Preliminary communication (original scientific work in progress, shorter than a regular scientific paper);
 - Disquisition and reviews on a certain topic based on scientific argumentation.
2. Scientific articles presenting experiences useful for advancement of professional practice.
3. Informative articles can be:
 - Introductory notes and commentaries;
 - Book reviews, reviews of computer programs, data bases, standards etc.
 - Scientific event, jubilees.

Papers classified as scientific must receive at least two positive reviews. The opinions of the Editorial Committee do not have to correspond to those expressed in the published papers. Papers cannot be reprinted nor published under a similar title or in a changed form.

ELEMENTS OF MANUSCRIPTS

For scientific or professional papers the following data should be provided:

1. Papers should not normally exceed 15 A4 pages, Times New Roman 12pt. For longer articles the authors should contact the journal editors.

2. Names and surnames of all authors should be written in the sequence in which they will appear in a published paper.
3. After each author's full name, without titles and degrees, an e-mail address should be specified as well as the full and official name of his or her affiliation. (For large organizations full hierarchy of names should be specified, top down).
4. The submission date should be provided.
5. The authors should suggest the category of their paper but the Editor-in-Chief is responsible for the final categorization.
6. An informative abstract not normally EXCEEDING 200 WORDS that concisely outlines the substance of the paper, presents the goal of the work and applied methods and states its principal conclusion, should accompany the paper. The abstract should be supplied in both languages used for publication. In the abstract, authors should use the terms that, being standard, are often used for indexing and information retrieval.
7. Authors should supply at least 3 but not more than 10 keywords separated by commas that designate main concepts presented in the paper. The list of keywords should be supplied in both languages used for publication.
8. If paper derives from a Master's thesis or Doctoral dissertation authors should give the title of the thesis or dissertation, as well as a date of its submission and names of responsible institutions.
9. If the paper presents the results of authors' participation in some project or program, authors should acknowledge the institution that financed the project in a special section "Acknowledgment" at the end of the article, before the "Reference" section. The same section should contain acknowledgment to individuals who helped in the production of the paper.
10. If the paper was presented at a Conference but not published in its Proceedings, this should also be stated in a separate note.
11. Authors can use footnotes, while endnotes are prohibited; however, too long footnotes should be avoided. Authors can add appendices to their paper.
12. The referenced material should be listed in the section "References" at the end of the paper. In the reference list authors should include all information necessary for locating the referenced work. All items referenced in the text should be listed here; nothing that was not referenced in the text should appear in this section.

EDITING CONVENTIONS FOR ACCEPTED PAPERS

1. Papers should be prepared and submitted using L^AT_EX (the journal style and all packages can be downloaded from the journal web site). Authors that are not familiar with L^AT_EX can prepare their papers using Word, as .doc, .docx, .rtf or .txt documents. These authors should not use any special formatting – the final formatting and transformation to L^AT_EX will be done by the Infotheca team.

2. The papers written in Serbian should use CYRILLIC alphabet because they will be printed in that script. The only exceptions are those parts of the text for which the use of the other script, such as Latin, is more appropriate. All scripts should be represented using Unicode encoding, UTF-8 representation.
3. Title of the paper should not be written in capital letters. The authors should keep the length of titles reasonable – preferably less than 90 characters. For all titles authors should provide a shorter title that will be used for page headers.
4. Italic type may be used to emphasize words in running text, while bold type or italic bold type can be used if necessary. Underlined text should be avoided. Please do not highlight whole sentences or paragraphs.
5. Paper can be divided in sections and subsections, but more than two levels of the section headings should be avoided. All sections and subsections will appropriately numbered. Appendices, if any, should come at the end of the paper and they will also be appropriately labeled. If using lists, do not use more than two levels of nesting.
6. All paragraphs should be separated by one empty line (one Enter).
7. Authors should avoid too wide tables keeping in mind that the journal is published on A5 paper and. All tables, illustrations, diagrams and photographs should not be wider than 72.5 mm (the width of one column) or (exceptionally) 150 mm (the width of the page). All illustrations should be prepared in some lossless format, for instance .png, .tif or .jpg and their resolution should be at least 300 dpi.
8. The authors are kindly requested to add (if possible) the link to the screen from which a screenshot was taken. When taking a screen shot of a part of some screen, authors are advised to use the Zoom possibility of the browser or other program. For diagrams that are produced with Excel, please provide the original .xls document.
9. All tables, illustrations, diagrams and photographs should be prepared as separate files, both in black-and-white for printing and in color for the on-line version. Captions that should be below tables, illustrations, diagrams or photographs should remain in the text. Each file should have the same name as the file containing the main text, followed by the type of material to which the ordinal number in the text is added. For instance, the file containing the fourth figure of the paper “Example” should be named Example_figure_4.
10. Please add additional document(s) that explain some specific aspects of formatting required for your paper, for instance, formulas prepared in L^AT_EX in a .pdf format.
11. URL addresses that appear in the paper should be placed in footnotes; the date when the site was visited should be given.

REFERENCES AND CITATION

1. Referenced material should be listed at the end of the text, within the unnumbered section References. The reference section should be complete; references should not be omitted. This section should not contain any bibliographic information not referenced in the main text. Referenced items should not be mentioned in footnotes.
2. Entries in the reference list should be ordered alphabetically by authors or editors names, or publishing organizations (when no authors are identified). If this list contains several entries by the same authors, these entries should be ordered chronologically.
3. For preparation of a reference list use Chicago Manual of Style reference list entry (www.chicagomanualofstyle.org).
4. Full names of journals, and not their short titles or acronyms, should be specified. Use the 10-point type for entries in the reference list.
5. All authors, whether they prepare their articles using L^AT_EX or Word, will prepare all the items from their References section using BibTeX templates that are given for all the examples at the Infotheca web site (<http://infoteka.bg.ac.rs/index.php/sr/upu-s-v-z-u-r>).