INFOtheca: **Journal of Digital Humanities**

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

Belgrade, Vol. 16, No. 1/2, August 2016

Impressum

JOURNAL REDACTION:
Journal Infotheca
+381 11 3370-211, infotheca@unilib.rs
11000 Belgrade, Bulevar kralja Aleksandra 71

PRINTED BY:
Dunav
Belgrade

Journal is published twice a year

Table of Contents

Application of a Structural Support Vector Machine Method to N-gram Based Text Classification in Serbian

UDC 811.163.41'322.2
DOI 10.18485/infotheca.2016.16.1_2.1

ABSTRACT: The paper presents classification results that were obtained using the Support Vector Machine method (SVM) over a hierarchically organized corpus of documents in Serbian. Two techniques derived from the SVM with structural output have been applied: multiclass flat classification and hierarchical classification. A common representation model of a document and a class or a hierarchy of classes the document belongs to, specific for this form of the SVM method, is based on different length byte n-grams. Four tf-idf statistics have been used that define significance of an n-gram for a specific document. The described techniques and statistics have been tested on a hierarchically structured subset of the Ebart corpus of newspaper texts. The results obtained for both types of classifiers are similar for the corpus as a whole, while hierarchical classifier performs better for most specific classes with a small number of texts.

KEYWORDS: hierarchical text classification, Support Vector Machine Method, Ebart corpus

PAPER SUBMITTED: 22 May 2015
PAPER ACCEPTED: 30 October 2015

Jovana Kovačević
jovana@matf.bg.ac.rs

Jelena Graovac
jgraovac@matf.bg.ac.rs

University of Belgrade
Faculty of Mathematics
Department for Computer Science and Informatics

1 Introduction

Text classification is one of the text mining tasks – an area of computational linguistics involving a set of techniques for extracting useful, hidden, previously unknown information from textual documents. In the case of text classification, hidden information is a class or a set of classes from a predefined set of classes that a text belongs to, based on its content (Manning and Schütze, 1999). Classification may

be performed manually, but such a process is time consuming and expensive. Availability of high speed computers made automatic classification the basis for efficient processing of large document collections and discovery of knowledge contained in them.

In automatic or semi-automatic text classification two different approaches are mainly used: lexical-semantic language resources based approach and machine learning based approach. Systems of the first type use lexical-semantic networks, such as WordNet (Miller, 1995) and FrameNet (Johnson et al.), along with resources and tools such as electronic dictionaries and lexicon grammars (Gross, 1997), semantic ontologies, named entity and proper names ontologies. These language resources provide for development of a classification model for a morphologically and derivationally exceptionally rich language as Serbian. Classifier is usually built manually based on rich morphological, syntactic and semantic information contained in lexical resources (Scott and Matwin, 1998) and it does not require existence of a set of pre-classified texts to be used for classifier training. As opposed to these, systems that follow another approach assume existence of pre-classified text corpora divided into training and test sets. Based on training data, classifiers (classification models), are built by using different statistical methods, e.g., Bayes classification (Eyheramendy et al., 2003), Conditional Random Fields (McCallum and Pereira, 2001), Hidden Markov Models (Yi and Beheshti, 2013), or methods based on Support Vector Machine model, neural networks (Sebastiani, 2002), nearest neighbors (Yang and Pedersen, 1997), decision trees (Quinlan, 1996), etc. Especially important is the multilingual EuroVoc classifier JEX (Steinberger et al., 2013), comprising of trained classifiers for 22 different European Union languages.

Choice of a method, as well as of an approach to a text classification problem, depends on two key factors: availability of language resources and availability of training data. If language resources such as lexicons, dictionaries and grammars and semantic networks exist, it makes sense to use the resources-based approach. This approach has to take into account characteristics and specificities of each language to be applied to, so in case of Serbian the fact that it uses two alphabets (Cyrillic and Latin), that orthography is phonologically based, morphological system rich, free word order in a sentence, agreement system very complex (Vitas et al., 2003). All these characteristics make preprocessing steps, such as feature selection and feature extraction, based on which classification is performed, quite complex. If corpora needed for training algorithms – in case of text classification those are classified text bases – exist or are easy and inexpensive to develop, it is convenient to apply the machine learning approach. In machine learning techniques, a classifier is generated automatically, by "learning" characteristics of classes based on a training dataset associated to each class. Data in these datasets are manually classified by domain

experts. After the training process, the classifier usually automatically generates a set of rules the data item has to satisfy in order to be classified in a specific class.

Depending on the number of classes, classification may be binary, when only two classes are defined or multiclass, when more than two possible classes are defined. Depending on whether classes may overlap or not, classification may be single-label, when a data item may be associated exactly one class, or multi-label, when a data item may be assigned one, zero or more than one class, i.e., classes may overlap. According to the structure defining relationship between classes, classification may be hierarchical or non-hierarchical. If classes are treated independently, without any structure defining relationships between them, it is a non-hierarchical classification. When the number of different classes, or the number of data items inside one class becomes very large, problems arise with accurate and efficient searching and managing of data at a class level. In that case classes are usually organized into tree-like structures and a hierarchical structure is introduced among them (Sun and Lim, 2001) (e.g., Yahoo hierarchy).

In (Graovac, 2014b) and (Pavlović-Lažetić and Graovac, 2010) a document classification method for Serbian is presented based on Serbian WordNet (Krstev et al., 2004) developed for Serbian by the Language Technology Group at the University of Belgrade, Faculty of Mathematics[1] and applied to the newspaper texts corpus Ebart[2]. Other language resources developed for Serbian by the Group have also been used – electronic dictionary (Vitas and Krstev, 2005), lexicon grammars (Vitas et al., 2003), proper names ontologies (Krstev et al., 2005).

In (Graovac et al., 2015; Graovac, 2014a; Graovac and Pavlović-Lažetić, 2014; Graovac, 2014b) machine learning – based classification methods are presented using n-gram method for text representation and k nearest neighbor method (kNN) for classifier development. Methods are language independent, applied to text corpora of the most widespread writings and languages, with different lexical, morphological, syntactic and orthographic characteristics (English, Chinese, Arabic, and Spanish), they are very simple and performed very well. In (Graovac, 2012), application of the machine learning n-gram method was presented to text classification in Serbian on the newspaper texts corpus Ebart. Classification is multiclass, multi-label and non-hierarchical. In this paper the structural support vector machine method is applied for the first time to text classification in Serbian (Ebart corpus). A hierarchical class structure has been defined over the flat corpus, and a sub-corpus has been extracted from the original corpus, consisting of documents corresponding, by content, to the classes in the hierarchical structure. Then two classification methods are applied to the hierarchically organized corpus, both inferred from the SVM with structured output (SSVM): multiclass (flat) classification (selection of one out of many classes

[1] www.matf.bg.ac.rs/∼cvetana/LT-pregled.html
[2] http://www.arhiv.rs/novinska-arhiva/

and hierarchical classification (selection of a class hierarchy the document belongs to). The common representation model of a document and a class or a class hierarchy the document belongs to, specific for this form of the SVM method, is based on different length byte n-grams.

Last, the model has been evaluated based on test data and accuracy of results has been determined by using one of the standard information retrieval measures – F1 measure combining recall and precision (Tan et al., 2006).

In what follows, the Ebart corpus will be presented first – the largest digital media documentation in Serbia, as well as the hierarchically organized sub-corpus extracted from the Ebart corpus for the purpose of testing the hierarchical classification methods (part 2).

Then the applied methodology will be sketched (part 3): SVM method with structured output and its adjustments for application to multiclass and hierarchical classification (section 3.1), byte n-gram concept (section 3.2) and a specific n-gram common representation of document and class (or class hierarchy) the document belongs to, as well as training and testing steps in application of this method (section 3.3). Evaluation measures are introduced in the section 3.4.

The main result of the paper – text classification result – will be presented in part 4. Evaluation results will also be presented as well as comparison with related (comparable) results. Finally, in part 5, we shall interpret and discuss obtained results, impact of applied methods and improvement possibilities.

2 Dataset

Ebart corpus represents the largest digital documentation of newspaper texts in contemporary Serbian language. It has been created in 2003 and up to nowadays it has warehoused more than 2,000,000 texts from the printed media. Last version of Ebart archive is classified according to theme units resembling the common columns in the newspapers: politics, foreign affairs, society, economy, chronicle, culture, fun, sports, media, feuilleton, readers' letters, and so on. In order to test structural support vector machines method on the problem of hierarchical text classification, we extracted a hierarchically organized sub-corpus of "flat" Ebart corpus and named it EbartHier. EbartHier includes all the articles from the daily newspaper "Politika", published in from 2003 to 2006, that belong to the following columns: Politics, Society, Economy and business, World economy and finances, Culture, Science and Technology, as well as all the articles from the "Sportski zurnal" magazine (published from 2003 to 2006) that belong to columns Basketball and Football. All the documents that belong to columns/classes that are similar by topic, have been grouped in the same class on the higher hierarchy level. In this way, classes Politics and society, Economy, Culture and science and Sport have been created. The obtained corpus

has a treelike structure represented in Figure 1. Classes are non-overlapping (one document can belong only to one class) and each document can be classified only in a class that is situated in the leaf of the hierarchy. Corpus is characterized by extremely uneven distribution of documents by classes (see Figure 1). The median of lengths of all the documents from the corpus is 327.75 words in a document (the shortest one has 7 and the longest one has 2576 words). Figure 2 represents a histogram of medians of documents' lengths (including the length of the shortest and the longest document) by all classes. All documents in the corpus are represented in Serbian Latin alphabet using the UTF-8 code scheme.

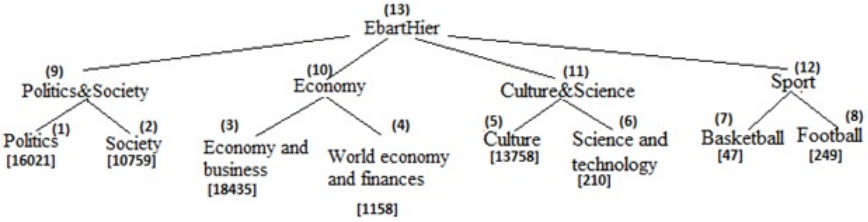


Figure 1. Treelike structure of EbartHier corpus. Each class name is assigned a number corresponding to that class. In the leaves, in square brackets, number of documents per class is displayed

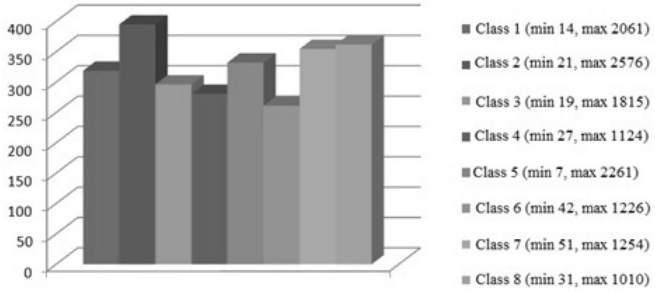


Figure 2. Histogram of mean lengths of documents in classes. Numbers in brackets correspond to the lengths of the shortest and the longest document in each class.

3 Methodology

3.1 Structural support vector machines method

Support Vector Machines method (SVM) has shown to be very efficient in text classification (Joachims, 1998). In this paper, we used Structural Support Vector Machines method (SSVM, (Tsochantaridis et al., 2004)) which represents generalization of the SVM method on structural output, for example array, tree, directed acyclic graph etc.

Before we represent the SSVM method, we will give a brief overview of the SVM method from the perspective of text classification. A standard approach in training predictors for binary classification is to learn the discriminant function $F(x)$ and to classify the input vector x according to the sign of the function $F(x)$. Since linear methods usually have efficient learning algorithms, it is common to assume that the function $F(x)$ is linear. Following this assumption, we can represent function $F(x)$ in the form $F(x) = \langle \omega, x \rangle$, where ω is the vector of learning parameters and „$\langle \dots \rangle$" is the notation for scalar product. Input vector x can be mapped into another space using the function Ψ and in that case, we write function $F(x)$ in the following way: $F(x) = \langle \omega, \Psi(x) \rangle$.

Binary classifier can predict whether a document belongs to a certain class or not, which means that its output can be -1, if the document does not belong to that class, or 1 otherwise. If we want to answer the question "Which class does the document belong to?" we need to turn to methods that predict structural output, namely SSVM. In the basic SVM method, output y is determined according to the sign of the discriminant function, that is $sgn(F(x)) = y$, where $y \in \{-1, 1\}$. It would be ideal if we could find analogous discriminant function $F(x)$ that maps the input dataset (in this case, corpus of documents) into the output dataset (in this case, classes). Since it is very difficult to create such function, we turn to the next best solution and create a function $F : X \times Y \to \mathbb{R}$ that measures how well the output y corresponds to the input x.

We would like to create the function F such that the larger value of $F(x, y)$, the better the output y corresponds to the input x, that is, in this specific case, given class better corresponds to the given text. In this generalization, the discriminant function becomes the function of two arguments, input and output, $F(x, y)$ where the output is not from the set $\{-1, 1\}$ yet it can represent array, tree, graph, etc. If we denote the set of all possible outputs by Y, no matter whether it contains arrays, graphs, trees or some other structure, SSVM predicts the output (class) that corresponds best to the input (text), i.e. it predicts the output vector y that maximizes the value of the function F for a given input vector x. To be more precise, SSVM predicts output based on the following equation: $y^* = \underset{y \in Y}{\operatorname{argmax}}(F(x^*, y))$. Analogously with

SVM, in SSVM we also assume that function F is linear in ω, as well as that the pair of vectors (x, y) can be more suitably represented by mapping into some other space using a function Ψ. Therefore, we can write function F as $F(x, y) = \langle \omega, \Psi(x, y) \rangle$.

Function Ψ represents a joint input-output vector for one input-output pair (x, y) and its form depends on the dataset that the method is being applied to. For example, one of applications of SSVM is creating ("predicting") a derivation tree in the given formal grammar and in that case, input vector x would represent a vector of words that appear in a sentence, output vector y would represent a derivation tree in the formal grammar and function $\Psi(x, y)$ would be a joint representation of the sentence and its derivation tree. This joint representation could be a vector whose dimension is equal to the total number of derivation rules, including rules of derivation of all words from the training dataset. Each element of this vector would correspond to one of all possible rules of the grammar, and its value would be equal to the total number of occurrences of the that rule. This representation is illustrated by the example in Figure 3.

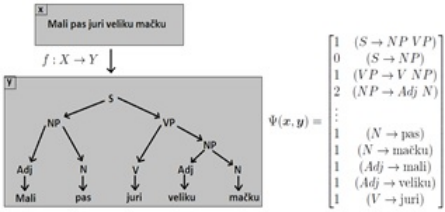


Figure 3. Example of joint representation of input vector x and output vector y, constructed according to the similar example of derivation in English grammar in (Tsochantaridis et al., 2004)

In the example shown in Figure 3, vector x consists of the words from the given sentence ("Mali pas juri veliku mačku", eng. "A small dog is chasing a big cat"), and vector y represents a derivation tree in the given formal grammar. Vector $\Psi(x, y)$ denotes that, in the sentence derivation, rule $S \to NPVP$ has been applied once, rule $S \to NP$ zero times, rule $VP \to VNP$ once, rule $NP \to AdjN$ two times and so on, rule $V \to juri$ once.

There are different formulations of the SSVM method (Tsochantaridis et al., 2005). In this paper, we used the so-called 1-slack margin rescaling formulation (Joachims et al., 2009). When a classifier is being trained, it is good for the margin

that separates the training examples to be as wide as possible, whereas on the other hand, wide margin can cause misclassification of some training examples. The 1-slack margin rescaling formulation learns parameters ω depending on a positive constant C that controls the trade-off between minimizing training error and maximizing the margin (this constant will be especially analyzed further in the paper). The described algorithm has polynomial complexity by the number of training examples, which has been proven in (Joachims et al., 2009).

3.2 N-grams

Let s be a string of symbols $s = s_1 s_2 ... s_N$ over alphabet Σ. An n-gram of the string s (for natural numbers n and N) is defined as any substring of adjacent symbols of the string s of length n. Totally $|\Sigma|^{\mathrm{n}}$ different n-grams can be defined over the alphabet Σ, where $|\Sigma|$ is its size (cardinality). N-grams represented in this way can be defined on word level, character level or byte level. For example, 2-grams (usually referred to as bigrams) over the string "dela, ne reči" on word level will contain only two bigrams "dela ne" and "ne reči", whereas character level bigrams (alphabet Σ is Serbian Latin) will be de; el; la; a,; ,_; _n; ne; e_; _r; re; eč; či. If the characters are coded by UTF-8 code scheme, letter "č" is coded by two bytes whose decade content is 196 141, respectively, character " " (space) is coded by one byte whose content is decade number 32 and so on. In that way, the entire string is written in the computer by an array of bytes with decade values 100 101 108 97 44 32 110 101 32 114 101 196 141 105. Thus, in case of languages over the Latin alphabet, n-grams on byte level and n-grams on character level are quite similar considering the fact that one character is mostly represented by one byte. The difference is usually also in the set of characters that is being considered (n-grams on the character level usually do not take into account the difference between big and small letters, punctuation symbols and digits), and the difference is especially significant when Cyrillic alphabet is used, or other alphabets like Arabic, Chinese etc. N-grams of bytes and n-grams of characters are equally used for text representation in solving different data mining tasks (Kešelj et al., 2003; Abou-Assaleh et al., 2004; Reddy and Pujari, 2006; Santos et al., 2011; Lui et al., 2014), with similar results. Although byte n-grams sometimes do not have specific meaning, especially for humans (for example, when they contain only one of two bytes that represent a character), their extraction from a text does not require the information about the used code scheme, which is why they represent simplified representation for computer processing. In this paper, we will use byte n-grams.

When used in natural language processing, some of the advantages of n-grams include relative insensitivity to spelling mistakes, the alphabet of symbols is known

in advance, independency of language and content, execution in one run, no need for any previous linguistic knowledge, and so on.

The basic problem with using n-grams is their exponential number with respect to the alphabet's cardinality. If Σ is English language alphabet and if we attach a space symbol to it, than $|\Sigma| = 27$. If we make a difference between capital and small letters and if we also include digits, than $|\Sigma| = 63$. It is clear that many algorithms with n-grams will be very expensive from the computational point of view even when $n = 5$ or $n = 6$ (for example, number of different 5-grams over the alphabet Σ is $63^5 \sim 10^9$).

Using models and techniques based on n-grams for natural language processing has shown to be an efficient approach. This approach has application in information retrieval (De Heer, 1974), text compression (Wiśniewski, 1987), detecting and correction of grammar mistakes (Zamora et al., 1981), detecting document's authorship (Kešelj et al., 2003) and so on.

3.3 Data representation

During this research, we developed two classifiers based on the original SSVM algorithm which have different representations of output data, in this case classes of documents. Both classifiers use the same representation of input data, namely byte level n-grams. Each position in n-gram vector is equal to tf-idf statistics' value for the given n-gram. Tf-idf statistics (tf-idf is short from term frequency-inverse document frequency) is usually defined in a way to reflect importance of an n-gram for a document in a corpus. This measure is directly proportional to the number of occurrences of the n-gram in the document, but it is also inversely proportional to the number of occurrences of the n-gram in the entire corpus. N-grams that have higher value of tf-idf statistics will be more significant for the document, more precisely, n-grams that occur more frequently in the document but are not that frequent in the entire corpus.

There are different variations for calculating values of tf and idf measures. In this paper, we used the following (Manning et al., 2008):

1. *classic tf-idf*: $tf \cdot \log\left(\frac{n}{n_k+1}\right)$
2. *log tf-idf*: $1 + \log(tf) \cdot \log\left(\frac{n}{n_k+1}\right)$
3. *boolean1 tf-idf*: $\log\left(\frac{n}{n_k+1}\right)$
4. *boolean2 tf-idf*: $\log\left(1 + \frac{n}{n_k}\right)$

where *tf* represents normalized frequency of n-grams in the corresponding document, n represents total number of documents in the entire corpus, and n_k represents number of documents in the corpus in which that n-gram occurs at least once.

Representation of output data, in this case classes, is different for two developed classifiers. Each document of EbartHier corpus is assigned one class situated in the leaf of the EbartHier hierarchy.

In the first classifier, each class is represented as a unique natural number, without considering the hierarchical relationship of the classes. If we enumerate the classes respectively as in Figure 1, set Y becomes equivalent to the set $\{1, 2, 3, 4, 5, 6, 7, 8\}$. In this way, we adjusted the basic SSVM method to work as a multiclass flat classifier.

In the first classifier, vector $\Psi(x, y)$, joint representation of input and output vectors, has dimension of $p \cdot q$, where p is the number of different n-grams, i.e. dimension of the vector x, and q is the number of different classes in the corpus. In that way, each class gets its block of size p in vector Ψ, which will contain zeros if the given document does not belong to the given class or values of the input vector x, otherwise. For example, if a document x belongs to class k, then the joint representation will be the following:

$$\Psi(x, y) = \left[\underbrace{0, \ldots, 0}_{\text{class 1}}, \ldots, \underbrace{x_1, \ldots, x_p}_{\text{class k}}, \ldots, \underbrace{0, \ldots, 0}_{\text{class q}} \right]$$

In the second classifier, each class is represented as a vector of nodes in EbartHier hierarchy. Each position in vector y corresponds to one node in the hierarchy, with value "1" if the node occurs in the path from the root to class' leaf and „0" otherwise. In the example in Figure 1, class economy and business would be represented as $(0, 0, 1, 0, 0, 0, 0, 0, 0, 1, 0, 0, 1)$. In this way, we adjusted the basic SSVM method to work as a hierarchical classifier.

In the second classifier, vector $\Psi(x, y)$ has dimension of $p \cdot r$, where p is the number of different n-grams, i.e. dimension of the vector x, and r is the number of nodes in the hierarchy. In that way, each node gets its block of size p in vector Ψ, which will contain zeros if the given document does not belong to the class that contains the given node or values of the input vector x, otherwise. For example, if document x belongs to class economy and business, represented as $(0, 0, 1, 0, 0, 0, 0, 0, 0, 1, 0, 0, 1)$, then the joint representation will be the following: $\Psi(x, y) = [0, 0, x, 0, 0, 0, 0, 0, 0, x, 0, 0, x]$ where 0 is the zero vector $0 = [\underbrace{0, \ldots, 0}_{\text{p times}}]$, and vector x corresponds to the input vector.

The basic difference in representations of vector y in flat and hierarchical classifier is that the flat one treats all classes individually, whereas the hierarchical one takes into account that the classes are part of a treelike hierarchy. This difference is reflected in the joint representation of input and output vectors, vector $\Psi(x, y)$, which has different dimensions in these two cases: in flat classifier, each position in vector Ψ is reserved for one class (which is situated in the hierarchy leaves) whereas

in hierarchical classifier each position is reserved for one node of the hierarchy, including leaves, inner nodes and root.

3.4 Implementation

Both classifiers have been constructed by adjusting publicly available and free SVM$^{\text{struct}}$[3] framework for SSVM method. SVM$^{\text{struct}}$ is available in several programming languages and for this purpose we used its implementation in programming language C. Adjustment of the basic implementation included adapting existing data structures for input vector x and output vector y, implementation of function which generates vector $\Psi(x, y)$ (joint representation of input vector x and output vector y), implementation of the loss function and implementation of the function for evaluating classifiers' quality.

The data have been prepared in the following way:

1. For each document, Text::Ngrams (Kešelj et al., 2003) tool has been used to generate the array of byte n-grams that the document contains as well as how many times each byte n-gram appears. Data for byte n-grams of length $\{2, 3, 4, 5, 6, 7\}$ have been generated.
2. We implemented a script that for a given document, based in its n-grams, generates input vector x in the specific format required by the SVM$^{\text{struct}}$ classifier

3.5 Experiments

We divided the EbartHier corpus of documents to training set and test set in proportion 2:1, which is one of the most frequent ways of division in text classification (Bellotti and Crook, 2009). From the entire corpus of 60,637 documents, the training set contains 40,426 documents and the test set 20,211 documents. Distribution of documents by classes also follows this proportion. It is displayed in Table 1.

Each document of the EbartHier corpus is represented by byte n-grams which determine the vector representation of the document suitable for the classifier, the so-called input vector x. Each byte n-gram from the corpus is assigned one position in the vector x and the value on the position is equal to the value of one of 4 tf-idf statistics for the corresponding byte n-gram. Depending on the length of byte n-gram ($n \in \{2 \dots 7\}$) and on the chosen tf-idf measure (classic, log, boolean1, boolean2), we generated 24 representations of the corpus ($n = 2$, measure=classic, ..., $n = 7$, measure=classic, $n = 2$, measure=log, ..., $n = 7$, measure=log, ..., $n = 2$, measure=boolean1, ..., $n = 7$, measure=boolean1, $n = 2$, measure=boolean2, ..., $n = 7$, measure=boolean2). In each representation, training sets and test sets consisted

[3] http://www.cs.cornell.edu/people/tj/svm_light/svm_struct.html

of the same documents so that the performance of classifiers could be compared on them.

In order to investigate whether a certain type of classifiers gives better results for certain length of byte n-grams or for certain tf-idf statistics, we performed flat and hierarchical classification for each corpus representation in the following way:

1. In order to find the optimal value of parameter C, we performed 10-cross validation on each training set.
2. For the value obtained, we have trained the classification model on the entire training set.
3. Testing and model evaluation have been performed on the test set.

At the end, we compared the performance of the trained classification models.

Class	Traning set	Test set	Entire corpus
1	10681	5340	16021
2	7173	3586	10759
3	12290	6145	18435
4	772	386	1158
5	9172	4586	13758
6	140	70	210
7	32	15	47
8	166	83	249
Total	40426	20211	60637

Table 1. Number of documents in the training set and in the test set by classes

3.6 Evaluation

In order to analyze performances of SSVM methods of hierarchical classification, we use the usual measures for quality of classification – precision (the percentage of correctly classified examples from all the examples assigned to a certain class), recall (number of test examples of the certain class that the classifier can recognize) and F-measure (F1), harmonic mean of precision and recall (Baeza-Yates et al., 1999):

$$Precision = \frac{TP}{TP + FP} \quad Recall = \frac{TP}{TP + FN}$$

$$F1 = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall}$$

where *TP* (True Positives) is the number of correctly positively classified documents, *TN* (True Negative) is the number of correctly negatively classified documents, *FP* (False Positive) is the number of incorrectly positively classified documents and *FN* (False Negative) is the number of incorrectly negatively classified documents.

These measures are defined for the case of binary classification (when there are only two classes). In case when there are more than two classes, it is necessary to find mean value of these measures by all classes. That can be achieved in two ways: by calculating macro-average, where each class is equally significant, or by calculating micro-average, where classes with larger number of documents are being favored. In macro-average, we first calculate measures for each class individually and then find mean value by the number of classes. In micro-average, we calculate values for *TP, TN, FP* and *FN* for each class individually and then we calculate values *TP, TN, FP* and *FN* by summing all *TP, TN, FP* and *FN* for all classes respectively. At the end, we calculate measures for the sums *TP, TN, FP* and *FN*. In this paper, we will use micro-average of F-measure (micro-F1). Basic disadvantage of these measures is that errors made on different levels of hierarchy are equally punished.

4 Results

We applied two variants of the SSVM method on the EbartHier corpus: in the first one, we performed flat classification, not taking into account hierarchical relationship of the classes, and in the second one we performed hierarchical classification. For every document from the corpus, 6 different n-gram representations have been generated, for byte n-grams of length from 2 to 7, for 4 different tf-idf measures presented in section 3.3. On each training set obtained in this way, we performed 10-cross validation which determined optimal value of the classifiers' parameter C, from the set of values from 10^{-2} до 10^{2}, with step 10. Figure 4 shows evaluation results for both types of classifiers, represented by micro-F1 measure, for corpus represented with byte n-grams of length from 2 to 7 and for all 4 tf-idf measures whereas Table 2 shows complete numerical data. Both types of classifiers (flat and hierarchical) show the same tendency that classifiers with classic measure have worse performances than others. Values of micro-F1 for other measures are approximate. Differences between two micro-F1 measures for different classifiers (flat and hierarchical) with the same measure are almost always under 1% (except for measure classic, where the differences are 2.32% and 1.66% for $n = 2$ and $n = 3$). Performances of the best classifiers for both types (flat and hierarchical) have been analyzed by classes. Among flat classifiers, classifier with measure boolean1 and byte n-gram length of 6 had the highest value of micro-F1 measure (90.43%) and among hierarchical classifiers, the

most successful was the classifier with measure boolean2 and byte n-gram length of 4 (micro-F1 is 90.17%). Considering that the number of documents by classes is different (displayed in Figure 1), classes can be divided into two groups: "large" (1, 2, 3 and 5) and "small" (4, 6, 7 and 8). Figure 5 shows the results of the analysis described. Results for two best classifiers are approximate for the "large" classes whereas differences are more distinguished for the "small" classes. Best hierarchical classifier is much better in predicting classes 6 and 7 than the best flat classifier. Both classifiers show highest accuracy for "small" class denoted by 8 (football).

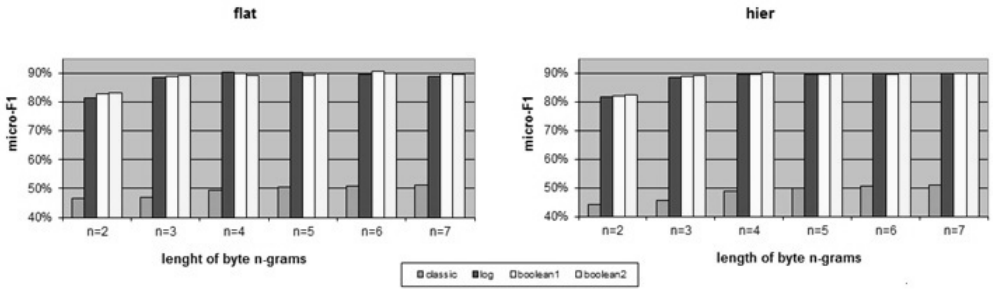


Figure 4. Results of the evaluation for two types of classifiers (flat and hierarchical) for different input data (byte n-gram's length from 2 to 7) and different measures

flat	classic	log	boolean1	boolean2
n=2	46.53%	81.46%	82.71%	83.25%
n=3	47.08%	88.27%	88.70%	89.28%
n=4	49.33%	90.05%	90.02%	89.18%
n=5	50.35%	90.05%	89.23%	89.77%
n=6	50.89%	89.60%	90.43%	89.74%
n=7	51.12%	88.88%	89.88%	89.60%

hier	classic	log	boolean1	boolean2
n=2	44.21%	81.75%	82.04%	82.45%
n=3	45.42%	88.34%	88.86%	89.24%
n=4	48.61%	89.39%	89.58%	90.17%
n=5	49.84%	89.46%	89.57%	89.70%
n=6	50.65%	89.84%	89.47%	89.78%
n=7	51.01%	89.88%	89.68%	89.69%

Table 2. Performances of flat (left table) and hierarchical (right table) classifier for different representations in input corpus, expressed by micro-F1 measure

5 Conclusion

Text classification of documents in Serbian language from a hierarchically organized corpus using the SSVM method shows approximate results for flat and hierarchical variants, for every tf-idf measure.

In total, the best result was achieved by the flat variant of the classifier for the boolean1 measure and its micro-F1 measure is 90.43%. This is a slightly better result than the previously published result for n-gram classification of a flat subset of the Ebart corpus where the micro-F1 measure was 88.5% (Graovac, 2012). Hierarchical classification model gives slightly better results than the flat one for some small classes.

Hierarchical classification presents weaker results than expected, which can be explained by the shallow hierarchy with a small number of documents, but also by the evaluation measure used. In evaluating classification results, especially hierarchical, it is not enough to count incorrectly classified examples, but it is also necessary to estimate their weight, i.e. the distance of the predicted class from the true class.

Having that in mind, we expect that by using better measure of evaluation, adapted for hierarchical classification, we could achieve results that outperform the results of the flat classification (Silla et al., 2011).

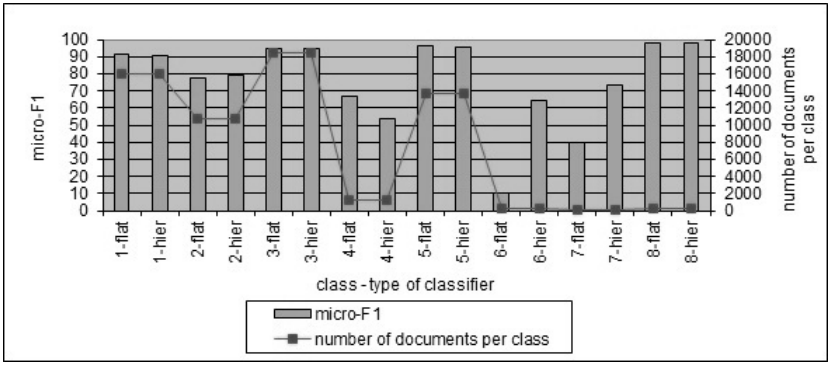


Figure 5. Results of the evaluation for the best classifiers for both types (flat and hierarchical). Number of documents in corpus per each class is also displayed.

References

Abou-Assaleh, Tony, Nick Cercone, Vlado Kešelj, and Ray Sweidan. "N-gram-based detection of new malicious code". In *Computer Software and Applications Conference*, COMPSAC 2004. Proceedings of the 28th Annual International, IEEE. Vol. 2 (2004), 41-42. Accessed September 1, 2015. `http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=1342667`.

Baeza-Yates, Ricardo, Ribeiro-Neto Berthier, et al. *Modern information retrieval*, vol. 463. ACM press New York, 1999.

Bellotti, Tony and Jonathan Crook. "Support vector machines for creditscoring and discovery of significant features". *Expert Systems with Applications*, vol. 36, no. 2 (2009): 3302-3308. Accessed September 1, 2015. `http://www.sciencedirect.com/science/article/pii/S0957417408000857`.

De Heer, T. "Experiments with syntactic traces in information retrieval". *Information Storage and Retrieval*, vol. 10, no. 3 (1974): 133–144. Accessed September 1, 2015. `http://www.sciencedirect.com/science/article/pii/0020027174900151`.

Eyheramendy, Susana, David D. Lewis, and David Madigan. "On the Naive Bayes model for text categorization". In *Proceedings of the Ninth International Workshop on Artificial Intelligence and Statistics Conference*, 332-339. Florida: Society for Artificial Intelligence and Statistics, 2003.

Graovac, Jelena. "Serbian text categorization using byte level n-grams". In *Proceedings of CLoBL 2012: Workshop on Computational Linguistics and Natural Language, 5th Balkan Conference in Informatics*, 93–97. 2012 Accessed September 1, 2015. `http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.416.6155&rep=rep1&type=pdf`.

Graovac, Jelena. "Wordnet-based text categorization technique". *INFOtheca – Journal of Information and Library Science*, vol. 14, no. 2 (2013): 2–17. Accessed September 1, 2015. `http://infoteka.unilib.rs/2013/br.2/eng/Infotheca-2-2013-Jelena-Graovac.pdf`.

Graovac, Jelena. "A variant of n-gram based language-independent text categorization". *Intelligent Data Analysis*, vol. 18, no. 4 (2014): 677–695.

Graovac, Jelena. "Text categorization using n-gram based language independent technique". Natural Language Processing for Serbian - Resources and Applications. In *Proceedings of the Conference "35th Anniversary of Computational Linguistics in Serbia"*, 124–135, 2014. Accessed September 1, 2015. `http://poincare.matf.bg.ac.rs/~jgraovac/publications/jg35CLS.pdf`.

Graovac, Jelena and Gordana Pavlović-Lažetić. "Language-Independent Sentiment Polarity Detection in Movie Reviews: A Case Study of English and Spanish". In *6th International Conference ICT Innovations 2014*, 13–22, 2014. Accessed September

1, 2015. http://poincare.matf.bg.ac.rs/~jgraovac/publications/jgSPD.pdf.

Graovac, Jelena and Jelena Kovačević, and Gordana Pavlović-Lažetić. "Language Independent n-Gram-Based Text Categorization with Weighting Factors: A Case Study". *JIDM - Journal of Information and Data Management*, vol. 6, no. 1 (2015): 4–17.

Gross, Maurice. "The construction of local grammars". In *Finite State Language Processing eds. Emmanuel Roche and Yves Schabs*, 329–354. Massachusetts: The MIT Press, 1997. Accessed September 1, 2015. https://halshs.archives-ouvertes.fr/halshs-00278316/document.

Joachims, Thorsten. "Text categorization with support vector machines: Learning with many relevant features". Springer Berlin Heidelberg, 1998.

Joachims, Thorsten, Thomas Finely, and Chun-Nam John Yu. "Cutting Plane Training of Structural SVMs". *Machine Learning Journal*, vol. 77, no. 1 (2009): 27–59. Accessed September 1, 2015. http://link.springer.com/article/10.1007/s10994-009-5108-8.

Johnson, Christopher R., Miriam R. L. Petruck, Collin F. Backer, Michael Ellsworth, Josef Ruppenhofer, and Charles J. Fillmore. "FrameNet: Theory and Practice", 2002. Accessed September 1, 2015. http://www.icsi.berkeley.edu/framenet.

Kešelj, Vlado, Fuchung Peng, Nick Cercone, and Calvin Thomas. "N-gram-based author profiles for authorship attribution". In *Proceedings of the conference pacific association for computational linguistics, PACLING*, vol. 3, 255–264. 2003. Accessed September 1, 2015. http://web.cs.dal.ca/~vlado/papers/pacling03.pdf.

Krstev, Cvetana, and Gordana Pavlović-Lažetić, and Ivan Obradović. "Using textual and lexical resources in developing Serbian wordnet". *Romanian Journal of Information Science and Technology*, vol. 7, no. 1–2 (2004): 147–161. Accessed September 1, 2015. http://xn--c1azn.xn--90a3ac/LicnePrezentacije/ivan_obradovic/Radovi/RJIS_2004.pdf.

Krstev, Cvetana, Duško Vitas, Denis Maurel, and Mickaël Tran. "Multilingual ontology of proper names". In *Proceedings of 2nd Language & Technology Conference*, April 21-23, 2005, Poznań, Poland, ed. Zygmunt Vetulani, 116–119. Poznań, Wydawnictwo Poznańskie, 2005. Accessed September 1, 2015. https://hal.archives-ouvertes.fr/hal-01108242/document.

Lui, Marco, Jey Han Lau, and Timothy Baldwin. "Automatic detection and language identification of multilingual documents". *Transactions of the Association for Computational Linguistics 2* (2014): 27–40. Accessed September 1, 2015. https://tacl2013.cs.columbia.edu/ojs/index.php/tacl/article/view/86/30.

Manning, Christopher, and Hinnrich Schütze. *Foundations of Statistical Natural Language Processing*. Cambridge: MIT Press, 1999.

Manning, Christopher, Prabhakar Raghavan, and Hinnrich Schütze. "Scoring, term weighting, and the vector space model". In *Introduction to Information Retrieval*. Cambridge: Cambridge University Press, 2008.

Lafferty, John, Andrew McCallum and Fernando C. N. Pereira. "Conditional random fields: Probabilistic models for segmenting and labeling sequence data". *Machine Learning*, 282–289. 2001

Miller, George A. "WordNet: a lexical database for English". *Communications of the ACM*,vol. 38, no. 11 (1995): 39–41. Accessed September 1, 2015. `http://nlp.cs.swarthmore.edu/~richardw/papers/miller1995-wordnet.pdf`.

Pavlović-Lažetić, Gordana and Jelena Graovac. "Ontology-driven conceptual document classification". In *Proceedings of the International Conference on Knowledge Discovery and Information Retrieval*, 383–386. 2010. Accessed September 1, 2015. `http://poincare.matf.bg.ac.rs/~jgraovac/publications/odcdc.pdf`.

Quinlan, J. R. "Improved use of continuous attributes in c4. 5". *Journal of Artificial Intelligence Research*, vol. 4 (1996): 77–90. Accessed September 1, 2015. `http://www.jair.org/media/279/live-279-1538-jair.pdf`.

Reddy, D. Krishna Sandeep, and Arun K. Pujari. "N-gram analysis for computer virus detection". *Journal in Computer Virology*, vol. 2, num. 3 (2006): 231–239. Accessed September 1, 2015. `http://link.springer.com/article/10.1007/s11416-006-0027-8`.

Santos, Igor, Javier Nieves, and Pablo G. Bringas. "Semi-supervised learning for unknown malware detection". In *International Symposium on Distributed Computing and Artificial Intelligence* vol. 91, 415-422. Berlin Heidelberg, 2011.

Scott, Sam, and Stan Matwin. "Text classification using wordnet hypernyms". In *Usage of WordNet in Natural Language Processing Systems: Proceedings of the Workshop*, 38-44, 1998. Accessed September 1, 2015. `http://www.aclweb.org/website/old_anthology/W/W98/W98-0706.pdf`.

Sebastiani, Fabrizio. "Machine learning in automated text categorization". *In ACM computing surveys (CSUR)*, vol. 34, no. 1 (2002): 1–47.

Silla J., Carlos N. and Alex A. Freitas. "A survey of hierarchical classification across different application domains". *Data Mining and Knowledge Discovery*, vol.22, no. 1–2 (2011): 31–72. Accessed September 1, 2015. `http://link.springer.com/article/10.1007/s10618-010-0175-9`.

Steinberger, Ralf, Mohamed Ebrahim, and Marco Turchi. "JRC EuroVoc Indexer JEX-A freely available multi-label categorisation tool". *arXiv preprint arXiv*, vol. 1309, no. 5223 (2013). Accessed September 1, 2015. `http://arxiv.org/ftp/arxiv/papers/1309/1309.5223.pdf`.

Sun, Aixin, and Ee-Peng Lim. "Hierarchical text classification and evaluation". In *Proceedings 2001 IEEE International Conference on Data Mining*, 521–528. Computer Society, 2001.

Tan, Pang-Ning, Michael Steinbach, and Vipin Kumar. *Introduction to data mining*, vol. 1. Pearson Education India, 2006.

Tsochantaridis, Ioannis, Thomas Hofman, Thorsten Joachims, and Yasemin Altun. "Support Vector Machine Learning for Interdependant and Structured Output Spaces". In *Proceedings of the twenty-first international conference on Machine learning*. 104. New York: ACM, 2004.

Tsochantaridis, Ioannis, Thomas Hofman, Thorsten Joachims, and Yasemin Altun. "Large Margin Methods for Structured and Interdependant Output Variables". *Journal of Machine Learning Research*,vol 6, (2005): 1453–1484. Accessed September 1, 2015. `http://machinelearning.wustl.edu/mlpapers/paper_files/TsochantaridisJHA05.pdf`.

Vitas, Duško and Cvetana Krstev. "Derivational morphology in an e-dictionary of Serbian". In *Proceedings of 2nd Language & Technology Conference*. April 21-23, 2005, Poznań, Poland, ed. Zygmunt Vetulani, 139-143. Poznań: Wydawnictwo Poznańskie, Accessed September 1, 2015. `http://poincare.matf.bg.ac.rs/~cvetana/biblio/ltc_134_vitas_2.pdf`.

Vitas, Duško , Cvetana Krstev, Ivan Obradović, Ljubomir Popović, and Gordana Pavlović-Lažetić. "An overview of resources and basic tools for processing of Serbian written texts". In *Proceedings of the Workshop on Balkan Language Resources, 1st Balkan Conference in Informatics, Thessaloniki, Greece*, 97–104, (2003). Accessed September 1, 2015. `http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.73.8096&rep=rep1&type=pdf`.

Wiśniewski, Janusz L. "Effective text compression with simultaneous digram and trigram encoding". *Journal of Information Science*, Vol. 13, No.3, (1987): 159–164.

Yang, Yiming, and Jan O. Pedersen. "A comparative study on feature selection in text categorization". In *Proceedings of the 14th International Conference on Machine Learning*, 412–420, 1997. Accessed September 1, 2015. `http://www.surdeanu.info/mihai/teaching/ista555-spring15/readings/yang97comparative.pdf`.

Yi, Kwan, and Jamshid Beheshti. "A text categorization model based on Hidden Markov models". In *Proceedings of the 31st Annual Conference of the Canadian Association for Information Science*, 275–287, 2013. Accessed September 1, 2015. `http://www.cais-acsi.ca/ojs/index.php/cais/article/view/420/585`.

Zamora, E. M., Joseph J. Pollock, and Antonio Zamora. "The use of trigram analysis for spelling error detection". *Information Processing & Management*,vol. 17, no. 6 (1981): 305–316. Accessed September 1, 2015. `http://www.sciencedirect.com/science/article/pii/0306457381900443`.

Ontology-based Recognition of Rhetorical Figures

UDC 811.163.41'322.2
DOI 10.18485/infotheca.2016.16.1_2.2

Miljana Mladenović
ml.miljana@gmail.com
University of Belgrade, Faculty of Mathematics

ABSTRACT: Automatic recognition of rhetorical figures (similes, irony, sarcasm, humor, metaphors, etc.) is increasingly used in natural language processing tasks, primarily to improve sentiment classification, machine translation, but also for analysis of linguistic structures on different levels. In this paper, it is proposed a method of automatic recognition and classification of rhetorical figures from the group of tropes that uses ontological inference rules in an ontology based on Serbian WordNet (SWN). A binary classification method was carried out on the rhetorical figure simile and evaluated by ROC curve ($AUC = 0.696$) which indicates that it can be successfully used in solving these types of tasks. It is also proposed a semi-automatic ontology learning method, for further learning of SWN ontology, by increasing the number and the type of relationships that can assist in the detection of figurative language in the texts in Serbian.
KEYWORDS: rhetorical figures, simile, ontology-based classification, WordNet

PAPER SUBMITTED: 25 February 2016
PAPER ACCEPTED: 09 May 2016

1 Introduction

Automatic recognition of rhetorical figures (figurative language) in text and its annotation are not new fields of natural language processing. One of the first pieces of research in the field of automatic identification and interpretation of rhetorical figures, which deals with the identification of *anaphora* (anaphora resolution), was given in the 1964 paper "A question-answering system for high school algebra word problems" (Bobrow, 1964), whereas in the paper "met*: A method for discriminating metaphor and metonymy by computer" (Fass, 1991) subjects of automatic

recognition were *metaphor* and *metonymy*. The literature describes methods of automatic recognition of *metaphors* (Mason, 2004; Hardie et al., 2007; Koller et al., 2008; Shutova et al., 2013), *anaphora* (Mitkov, 2002; Poesio and Artstein, 2010), *metonymy* (Farkas et al., 2007; Leveling, 2007; Nicolae et al., 2007), *irony* (Carvalho et al., 2009; Veale, 2012), *sarcasm* (Tsur et al., 2010; González-Ibáñez et al., 2011) etc. However, basic sentiment analysis (SA) methods mainly explore literal meaning of the text not taking into account figurative language.

In (Hao and Veale, 2010) authors analyzed figurative language and showed that certain rhetorical figures (e.g. *irony*) in text act as sentiment polarity *modifiers*, whose role in changing the sentiment polarity of words or phrases appears in the scope of the role of *valence shifters* presented in the paper "Sentiment Classification of Movie Reviews Using Contextual Valence Shifters" (Kennedy and Inkpen, 2006). In general, the *valence shifters* are words and phrases that reduce, increase or completely change the polarity of emotion words or phrases that they appear with (e.g. sentiment polarity of the word *happy* is changed by the phrase *not at all*). A certain group of rhetorical figures, which can affect the sentiment polarity of words or phrases in text, act in a similar way. By definition, figures *irony* and *sarcasm* change polarity, *dysphemism* and *hyperbole* enhance the existing level of sentiment expression, while *litotes* and *euphemism* reduce that level. *Metaphor*, *metonymy*, *oxymoron*, and *simile* have more complex mechanisms of action in both directions of changing the sentiment strength and polarity.

The process of automatic or semi-automatic identification of rhetorical figures in text can improve the sentiment analysis, machine translation and other natural language processing tasks. For example, the traditional sentiment polarity classification system (Pang et al., 2002) would interpret sentences "*He is quick as a rabbit*" and "*He is quick as a snail*" in the same way. The inclusion of figurative language recognition would show that the second sentence can be noted as ironical, which would invert its sentiment polarity compared to the literal meaning. Research published in (Reyes and Rosso, 2012b) has shown that the accuracy of sentiment classification can be significantly improved (from 54% to a maximum of 89.05%) when predictors of identifying figurative language were included, compared to a set of predictors that treated the text in the literal sense. In (Rentoumi et al., 2010) a sentiment analysis task has been improved by integration of a machine learning (ML) method with the method based on the rules which recognize the use of figurative language. The paper (Williams et al., 2015) discussed the role of idioms in sentiment analysis. In two experiments, the authors showed improvement of classification, measured by F-measure (from 45% to 64%, and from 46% to 61%), when idioms that carry figurative meaning were used as predictors of classification. In the paper (Carvalho et al., 2011) results of the sentiment analysis on comments published on the web portals of newspaper show that 11% of the observed set of comments would have

been incorrectly labeled as positive, had the analysis and identification of rhetorical figure *irony* not been used. Therefore, sentiment analysis of texts using figurative language is a new challenge in the field of natural language processing. In 2015, for the first time on a global scale, a task of this kind was set — sentiment analysis of figurative language in Twitter (International Workshop on Semantic Evaluation — SemEval-2015).[1]

Research in the field of processing of figurative language moves in two directions:

- to improve sentiment classification methods (Reyes and Rosso, 2012b; Rentoumi et al., 2010);
- to better understand the structure of a language — in (Veale and Hao, 2009) comparative analysis of ironic similes in English and Chinese languages showed that *irony* is a linguistic and also a cultural phenomenon, because a set of ironic simile expressions used in the paper as similes that are used in English can be applied in the Chinese language only in amount of 3–4%.

In the study of figurative language semantic networks such as WordNet (Mason, 2004; Barbieri et al., 2015), ontology (Harris and Di Marco, 2009; Kelly et al., 2010), lexical resources such as corpora (Mason, 2004; Hao and Veale, 2010; Reyes and Rosso, 2012a), specialized dictionaries of emoticons and punctuation (Carvalho et al., 2009; González-Ibáñez et al., 2011; Barbieri et al., 2015) and lexicons like SentiWordNet (Rentoumi et al., 2010; González-Ibáñez et al., 2011; Barbieri et al., 2015) play an important role.

In this paper we propose a method of automatic recognition of rhetorical figures belonging to the group of tropes, using the rules defined in an ontology SWN which is based on the Serbian semantic network WordNet. Afterwards, we propose an ontology learning method in terms of increasing the number and type of relations in SWN ontology that can help in the detection of such figures in texts in Serbian.

2 Methods for automatic recognition of rhetorical figures

Automatic recognition of rhetorical figures in text depends on the nature of a figure and its language structure. In order to build an effective system of automatic recognition of figurative language, it is necessary, first of all, to formally define and describe the rhetorical figures. The process of building the first formal domain ontology of rhetorical figures for Serbian (*RetFig*) is shown in the paper "Ontology of Rhetorical Figures for Serbian" (Mladenović and Mitrović, 2013). The *RetFig* ontology describes 98 rhetorical figures. For each of them, rhetorical and linguistic group

[1] http://alt.qcri.org/semeval2015/index.php?id=tasks (Retrieved on March, 23th 2016.)

they belong to were defined, the linguistic scope, objects and elements, mutual relationship of objects and elements, as well as linguistic operations involved in the process of their creation. All figures are divided into four groups: figures of pronunciation, figures of construction, figures of twisted meaning — tropes and figures of thought. Figures of pronunciation are based on the influence of certain letters in text (phonemes in speech). The repetition of a certain letter or group of letters in text (phonemes or groups of phonemes in speech), their omission or insertion in unexpected places, imitating certain sounds and noises from nature, affects the increase or decreases the importance of linguistic structures over which they run. This type of figures has no effect on the meaning of text and does not change it, but only emphasizes its basic meaning. Figures of construction are formed by changing the usual arrangement of words in a sentence or in another larger text portion (verse, for example). This kind of figures also does not change the basic meaning of the language structures that are being built. In contrast to these groups, figures of twisted meaning — tropes and figures of thought change the basic meaning of a word or larger text portion in their scope.

Automatic recognition of rhetorical figures of pronunciation and construction in text can be carried out using regular expressions as applied to the works (Gawryjolek et al., 2009; Hromada, 2011). For example, the figure *antimetabole*, belonging to the group of figures of construction, is defined by the repetition of a word or phrase which stands in the first part of the sentence and also in the second, but in reverse order.[2] The paper (Gawryjolek et al., 2009) describes a model of automatic recognition of such kind of figures, as well as the appropriate tool for automatic annotation (Java Annotation Tool of Rhetoric — Jantor). It gives the syntax pattern for automatic recognition of *antimetabole* in the form $[W_a] \ldots [W_b] \ldots [W_b] \ldots [W_a]$ in which it finds the occurrence of this figure in a sentence, where W_a and W_b are words or phrases in the sentence which in the second part of the sentence appear in reverse order compared to the first. Results of Jantor show that the rhetorical figures of pronunciation and construction can be successfully recognized, because it is possible to define regular expressions that define them unambiguously. Both classes of figures are based on syntactic and morphological operations within the grammatical rules of a natural language. However, the structure of rhetorical figures in classes of figures of twisted meaning — tropes and figures of thought cannot be defined and recognized by syntactic or morphological patterns.

In processes of automatic recognition and annotation of rhetorical figures that can change meaning of the text ML methods are successfully used. In the works that dealt with the *metonymy* recognition, supervised ML methods, such as making lists (Markert and Nissim, 2002), the maximum entropy (Farkas et al., 2007), k-

[2] For example: "*We eat to live, not live to eat.*", "*Write as you speak, speak as you write.*"

nearest neighbors (Leveling, 2007)[3], logistic regression (Nicolae et al., 2007) are used. In the paper (Tsur et al., 2010) a method of semi-supervised learning to identify and classify sarcastic tweets and comments from the site e-commerce Amazon is proposed. Support vector machines and logistic regression were used in the process of classifying sarcastic tweets and both positively and negatively sentiment polarized tweets without *sarcasm* (González-Ibáñez et al., 2011).

Besides by ML methods, *irony* and *sarcasm* can be successfully detected and by other techniques. In (Carvalho et al., 2009) authors created eight forms (e.g. $P_{laugh} = (LOL|AH|EMO+)$, $P_{quote} = (ADJ_{pos}|N_{pos})\{1,2\}$) which indicate the existence of structural *irony* in comments published on news portals in Portuguese language. The crowdsourcing method proposed by (Filatova, 2012) is used in the detection of *irony* and *sarcasm* in Amazon product reviews, while author (Veale, 2012) defined automatic extraction method of semantic knowledge in examples of using the *simile* figure in order to discover examples of ironic comparisons. In (Mitrović et al., 2015; Mladenović et al., 2016a) authors used a crowdsourcing method for detecting *similes* comparing it with the algorithm proposed for automatic extraction of simile candidates based on frequency of occurrence in an annotated corpus.

In this paper, in the process of detection of rhetorical figures belonging to the group of tropes we propose usage of ontological reasoning of the ontology derived from the semantic network Serbian WordNet (SWN).[4]

3 Ontology-based recognition of rhetorical figures tropes

Ontology is one of the forms of knowledge representation. If an ontology is described by a formal language and stored in a computer-readable format, it is a formal ontology. Understood as a "specification of shared conceptualization" (Gruber, 1993), ontology indicates the kind of knowledge that can be transferred, exchanged and used. According to (Devedžić, 2010), the main purpose of ontology is to be shared and reused by various intelligent agents and applications. Depending on which part of reality it describes, an ontology can be:

- top level ontology — when it describes the general concepts and the knowledge that it represents is comprehensive, systematized and applicable in a wide range of applications;
- domain ontology — when it represents knowledge of a certain domain;
- task ontology or application ontology — contains only the knowledge necessary to carry out the given class tasks.

[3] A software package TiMBL, based on the implementation of the algorithm Memory-Based Learning which is derived from the method "k-nearest neighbor", is used.

[4] http://sm.jerteh.rs/MemberZone/eW3.aspx?id=miljanam

In this paper, we propose a method for recognition of rhetorical figures that belong to the class of tropes, which, by definition, are those figures whose role is based on changing the basic meaning of words or phrases within a scope represented by surrounding words, phrase, verse or a sentence. Tropes base their role on the following characteristics:

- replace the meanings of a word, substituting that word with another word or phrase (e.g. in the case of an utterance representing *irony* "*You're my best friend*", the word *worst* is substituted by the word *best*);
- assign multiple meanings to a word or a phrase (e.g. in the case of an utterance representing figure of *synecdoche* "*He got a roof over their heads*" a phrase *roof over their heads* can get the meaning of words: *home*, *house*, *safety*, *shelter* or their synonyms);
- generation of a new meaning of a word or a phrase (e.g. figure *oxymoron* combines the concepts of opposite meaning in the new term, for example, *virtual reality*).

In the case of tropes it is not possible to define the morphological or syntactic patterns for recognitions, but it is possible to determine what kind of semantic relations exists between participants who build a certain figure and on that basis to define rules of inference in an ontology obtained from the semantic network WordNet (Fellbaum, 1998). Using a SPARQL query against a domain ontology *RetFig* (Mladenović and Mitrović, 2013) we can get all tropes used in Serbian.

```
select distinct ?figura
where { ?figura ont:jeNadObjektom ?objekt.
?figura ont:jeRetorickaGrupa ?retGrupa.
?objekt ont:naziv ?nazivObjekta.
?retGrupa ont:naziv ?nazivRetGrupe.
FILTER (?nazivObjekta ="REC"  &&
?nazivRetGrupe="FIGURE_TROPI")}
order by ?figura
```

The SPARQL query produced a collection of 26 rhetorical figures (Figure 1), which belong to the group of figures tropes (?nazivRetGrupe="FIGURE_TROPI") and are characterized by some form of change in the usual meaning of a word (?nazivObjekta="REC"). At the top of this part of the *RetFig* ontology is a class named Tropi (tropes). It contains three subclasses (VišestrukoZnačenje, IzmenjenoZnačenje, NovoZnačenje) which define the way in which is changing the usual meaning of the word on which a figure is applied (multiple meaning, changed meaning and a new meaning). The idea is, if it is possible, for each of these figures to be defined by rules (expressed using the Semantic Web Rule Language — SWRL)[5] over ontology SWN.

[5] Semantic Web Rule Language (SWRL) is the language used in Semantic Web for presenting formal logical expression, obtained by combining the features of OWL DL language

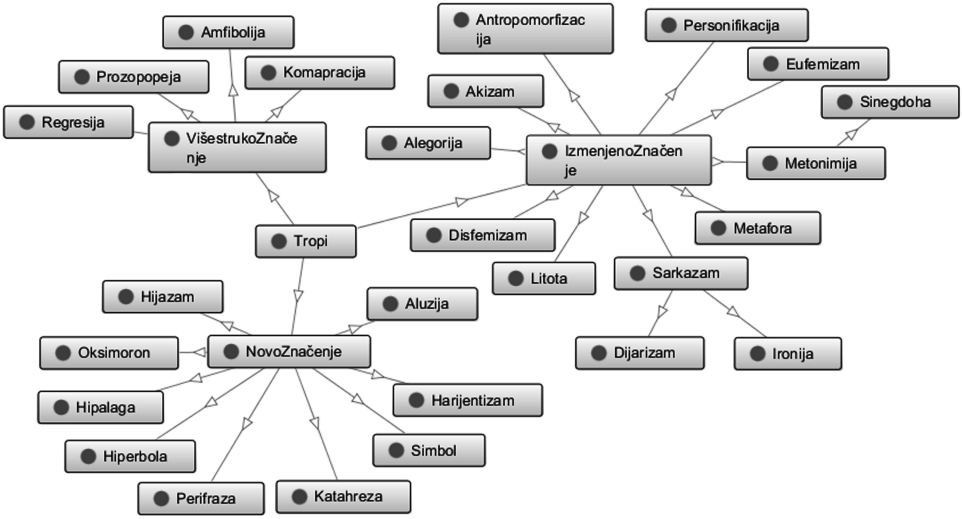


Figure 1: A part of taxonomic hierarchy of classes in an ontology *RetFig* concerning the figures of tropes

Explaining the case of figure *irony*, we will explain rules in the SWN ontology by which an occurrence of the corresponding figure in the text can be recognized. Verbal *irony* is the usage of words or phrases to say something opposite to the true meaning. The true meaning of the word is disguised and is opposite to the meaning of the used word or phrase. Examples of verbal irony are based on several types of semantic relations (Bagić, 2012), and commonly used forms in Serbian language based on the usage of an adjective instead its opposite (antonym) adjective, as in the examples:

– "*He's just brilliant!*" ("*The hidden meaning of the claim that someone is really stupid.*")
– "*See how skinny he is!*" ("*Hidden claim that someone is fat.*")

Another form consists of a noun and an adjective, but the adjective whose hidden meaning is opposite to the meaning of an adjective that is commonly associated with that noun.

– "*Fast as a turtle.*"

(Web Ontology Language for Description Logic) and RML language (Rule Markup Language).

– "*Brave as a rabbit!*"

The paper (Veale and Hao, 2009) used the term "*ironic comparisons*" to describe a specific type of *irony* derived from the figure simile. In previous examples, the adjective *fast* has a hidden meaning, the meaning of its antonym *slow* which is a natural feature of a noun *turtle*, while the adjective *brave* has a hidden meaning of the adjective *fearful*, which is a natural feature of a noun *rabbit*. Semantic relationships that exist in the ontology SWN and which can be used for generating candidates for the detection of figure *irony* are relations *specificOf* / *specifiedBy* and a relation *near_ antonym*. Relations *specificOf* / *specifiedBy* represent a pair of inverse relations (Mladenović et al., 2016a) which link an instance of a noun synset class with that instance of an adjective synset class which represents natural feature (is specific of) of the noun instance. The relation *near_ antonym* links two instances of an adjective synsets class which are direct antonyms mutually. Therefore, the rule for generating implicit relation representing *irony* is a relation between an instance of a noun synset class and an instance of adjective synset class. In ontology SWN it can be expressed as an OWL rule

$$\{?n : specifiedBy\ ?p1.\ ?p1 : near\_antonym\ ?p2\} \Rightarrow \{?n : Irony\ ?p2\}$$

or as SWRL rule

$$specifiedBy(?n, ?p1), near\_antonym(?p1, ?p2) \rightarrow Irony(?n, ?p2) \quad (1)$$

The result of inferencing by the rule (1) is equivalent to the result obtained by SPARQL query (Figure 2) in the ontology SWN.

Query result gives the candidates of figure *irony* in the form of a pair (adjective2, noun), for example, (*hrabar-brave*, *zec-rabbit*),(*brz-fast*, *kornjača-turtle*), (*brz-fast*, *puž-snail*), (*spor-slow*, *strela-arrow*), (*spor-slow*, *ideja-brainstorm*), (*lak-lightweight*, *slon-elephant*), (*vruć-hot*, *kamen-stone*), but also gives candidates of figure *simile* in the form of a pair (adjective1, noun), e.g. (*fearful*, *rabbit*), which is equivalent to OWL rule

$$\{?n : specifiedBy\ ?p1\} \Rightarrow \{?n : Simile\ ?p1\}$$

or

$$specifiedBy(?n, ?p1) \rightarrow Simile(?n, ?p1) \quad (2)$$

Query result in Figure 2 is analogous to the result of reasoning based on rules (1), but it can be observed that in natural language is not common to use some of the candidates for generating figure *irony*. Based on intuition, which at this point is not proven, it can be said that natural candidates for figure *irony* are

SPARQL query:

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX swn30: <http://www.mmiljana.com/swn30#>
SELECT ?adjective1 ?adjective2 ?noun
        WHERE { ?adjective1 swn30:near_antonym ?adjective2.
                ?adjective2 swn30:specificOf ?noun.
}
order by ?adjective1
```

adjective1	adjective2	noun
hrabar	plašljiv	zec
brz	spor	kornjača
brz	spor	puž
spor	brz	strela
spor	brz	ideja
lak	težak	slon
vruć	hladan	kamen

Figure 2: Examples of figures *irony* and *simile* obtained by SPARQL query in ontology SWN

those in which the sentiment polarity of the subject (?adjective1) in RDF triplets[6] (?adjective1 swn30:*near_antonym* ?adjective2) is positive. For example, more natural examples of irony are claims *fast as a turtle* or *lightweight as an elephant* than *slow as arrow* or *heavy as a feather*. Therefore, the query can be enhanced with an additional condition that take into account only those RDF triples in which the value of negative sentiment polarity of the subject is zero. Since each synset in the Serbian WordNet is labeled by a rate of positive and negative sentiment polarity (Mladenović and Mitrović, 2014) using the resource SentiWordNet (Baccianella et al., 2010), the SPARQL query can be extended to control the data property swn30:sentimentNegative and selected (filter) only those elements in which the value of that property is zero (Figure 3).

SWRL rule which includes an additional condition shown in Figure 3, is given below

$$\begin{aligned} specifiedBy(?n, ?p1), near\_antonym(?p1, ?p2), \\ sentiment\_negative(?p1, ?sent), swrlb : equal(?sent, 0) \ \rightarrow Irony(?n, ?p2) \end{aligned} \tag{3}$$

[6] RDF is a data model that uses a form of triplets (subject-predicate-object) to describe the semantic web resources, provide storage of data in graph databases and knowledge representation in ontological models.

SPARQL query:

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX swn30: <http://www.mmiljana.com/swn30#>
SELECT ?adjective1 ?adjective2 ?noun
        WHERE { ?adjective1 swn30:near_antonym ?adjective2.
                ?adjective2 swn30:specificOf ?noun.
                ?adjective1  swn30:sentimentNegative ?sent.
                FILTER regex( str(?sent), "0,00000")
}
        ORDER by ?adjective1
```

adjective1	adjective2	noun
hrabar	plašljiv	zec
brz	spor	kornjača
brz	spor	puž
vruć	hladan	kamen

Figure 3: Improving search for examples of figure *irony*

Let us now consider the structure of rhetorical figure *oxymoron* which, by definition, represents merging of terms with opposite meanings into a new concept. Examples of *oxymoron* found in SWN ontology are:
primerRecvirtualna stvarnost (*virtual reality*), *glasna tišina* (*loud silence*), *luda pamet* (*crazy clever*), *živi fosil* (*living fossil*), *vatreni led* (*fire-ice*), etc. The rules for generating candidates for figure *oxymoron* in ontology SWN can be expressed as

$$\{?p1 : near\_antonym\ ?p2.\ ?p2 : derived\text{-}pos\ ?n\} \Rightarrow \{?p1 : oxymoron\ ?n\}$$
$$\{?p1 : near\_antonym\ ?p2.\ ?p2 : be\_in\_state\ ?n\} \Rightarrow \{?p1 : oxymoron\ ?n\}$$

or as SWRL rules:

$$near\_antonym(?p1, ?p2),\ derived\text{-}pos(?p2, ?n) \rightarrow oxymoron(?p1, ?n) \quad (4)$$
$$near\_antonym(?p1, ?p2),\ be\_in\_state(?p2, ?n) \rightarrow oxymoron(?p1, ?n) \quad (5)$$

The rules (4) and (5) can also be obtained by SPARQL query (Figure 4) in the SWN ontology. Query result gives candidates for generating an *oxymoron* figure in the form of a pair (adjective1, noun), for example (*glasan-loud*, *tišina-silence*). At Figure 4 we can see that RDF triples (adjective1, adjective2,noun) containing adjectives that are mutual antonyms become candidates for the instantiation of this figure when one of adjectives is in relationship by any of relations *derived-pos* and *be_in_state* with the corresponding noun. In SWN, relations *derived-pos* and *be_in_state* are lexical relations between synsets belonging to different part-of-speech

(cross-part of speech relations — XPoS), in this case between adjectives and nouns (Koeva et al., 2008).

SPARQL query:

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX swn30: <http://www.mmiljana.com/swn30#>
SELECT ?adjective1 ?adjective2  ?noun
        WHERE {
        ?adjective1 swn30:near_antonym ?adjective2.
        {
            { ?adjective2 swn30:derived-pos ?noun.      }
             UNION
            { ?adjective2 swn30:be_in_state ?noun.      }
        }
        }
```

adjective1	adjective2	noun
ružan	lep	lepota
virtualan	stvaran	stvarnost
spor	brz	brzina
vruć	hladan	temperatura
glasan	tih	tišina

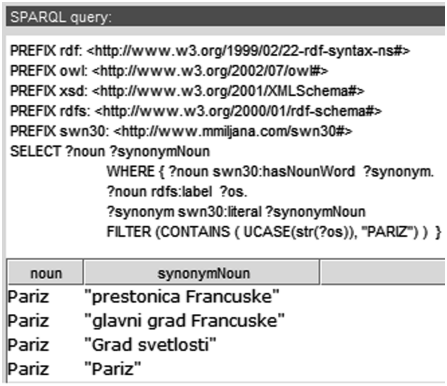
Figure 4: Examples of a figure oxymoron obtained by SPARQL query in ontology SWN

Beside the lexical-semantic relations, as an indicator of the existence of certain figures of tropes, an implicit relation *synonymy* can also be used. For example, we can find candidates for the figure *periphrasis* (*circumlocutions*) (Mitrović, 2014), when the term is described or replaced with more words using some essential properties of this concept (Figure 5). So the word *Paris* can be replaced by its synonyms existing in SWN: *Grad svetlosti* (*The City of Lights*), *Prestonica Francuske*, *Glavni grad Francuske*[7] (*The capital of France*), so pairs like (*Pariz — Grad svetlosti — The City of Lights*) become candidates for the instancing of this figure.

4 SWN Ontology Learning

In the previous section we described the rules defined in the SWN ontology which automatically generate candidates for some rhetorical figures belonging to group of figures tropes. However, the main problem in engineering an ontology, according to (Devedžić, 2010), is the ability of changing the ontology simultaneously with

[7] All examples extracted from SWN are presented in its original form, in Latin.



SPARQL query:

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX swn30: <http://www.mmiljana.com/swn30#>
SELECT ?noun ?synonymNoun
        WHERE { ?noun swn30:hasNounWord  ?synonym.
        ?noun rdfs:label  ?os.
        ?synonym swn30:literal ?synonymNoun
        FILTER (CONTAINS ( UCASE(str(?os)), "PARIZ") ) }
```

noun	synonymNoun
Pariz	"prestonica Francuske"
Pariz	"glavni grad Francuske"
Pariz	"Grad svetlosti"
Pariz	"Pariz"

Figure 5: Examples of the figure *periphrasis* obtained by SPARQL query in SWN ontology

changing the knowledge in the domain. Ontology learning using automatic and semi-automatic methods of extraction and annotating new knowledge in the domain and integrating it with existing knowledge in the ontology, largely depends on the tools that are applied for that purpose. Since we use ontology SWN to recognize different rhetorical figures, it is very important to learn different forms of figurative speech from examples used in a natural language. Since that recognition, as we described in previous section, depends on relations in the SWN ontology (*specificOf*, *specifiedBy*, *near_ antonym*, *derived-pos*, *be_ in_ state* and *synonym*) and entities that are linked by them, by ensuring automatic or semi-automatic intensive and continuous increase of the number of these relations and linked entities will get the ontology to be more complete and therefore more effective in the task of recognition. Diagram of the SWN ontology learning is shown in Figure 6.

SWN ontology learning, based on a method of automatic enlargement of semantic network Serbian WordNet, is proposed in (Mladenović et al., 2016a). The proposed method uses the annotated part of the digital Corpus of the contemporary Serbian language[8] (Utvić, 2014) to generate relations *specificOf* and *specifiedBy* between synsets existing in SWN. In this paper we propose a semi-automated method of enlargement of SWN which uses Web tools to facilitate the connection of synsets using some of the lexical-semantic relations defined in the SWN. The user interface is shown in Figure 7 and represents a tool that can link two synsets, previously chosen by a user, by one of the available relations. On the left side of Figure 7 for the entered word (e.g. lep — beautiful), underneath appear all the meanings of the

[8] http://korpus.matf.bg.ac.rs/prezentacija/korpus.html

Figure 6: SWN ontology learning from Corpus of contemporary Serbian language

word that are defined in SWN. On the right side of the same figure for the entered word (e.g. slika — picture), also below, all its meanings defined in SWN are shown. The user selects the appropriate meaning of each of the entered word and connects them by choosing a relation from a list of relations given in the middle column. If the user selects a relation for which there is an inverse one (e.g. *SpecificOf-SpecifiedBy*) it is automatically generated and the corresponding inverse relation between synsets of concepts that are selected according to meaning.

5 Evaluation of ontology-based method for recognition of rhetorical figure *simile*

In order to evaluate recognition of the rhetorical figure *simile* in texts in Serbian, we carried out testing using the collection of various texts consisting of 10 digitized writings of various genres: children's songs, fairy tales, comedies, novels and essays, all listed in Appendix A. To this collection were added two datasets (named *news* and *films*) of articles which are used in (Mladenović et al., 2016b) as testing sets in sentiment analysis in Serbian, composed of film reviews and news from news portals in Serbian language (bars labeled as *vesti*-news and *filmovi*-movies in Figure 8). We built a tool that, with the help of regular expressions, parses the input text and extract linguistic structures in the form of candidates to identify a simile. We used regular expressions "РЕЧ *као* РЕЧ {1,2}", "РЕЧ *nonym* РЕЧ {1,2}" (WORD *as* WORD {1,2}). From the extracted examples, we manually removed all those who are not in the form "РЕЧ *као* ИМЕНИЦА" (WORD *as* NOUN), "РЕЧ *као* ПРИДЕВ ИМЕНИЦА" and

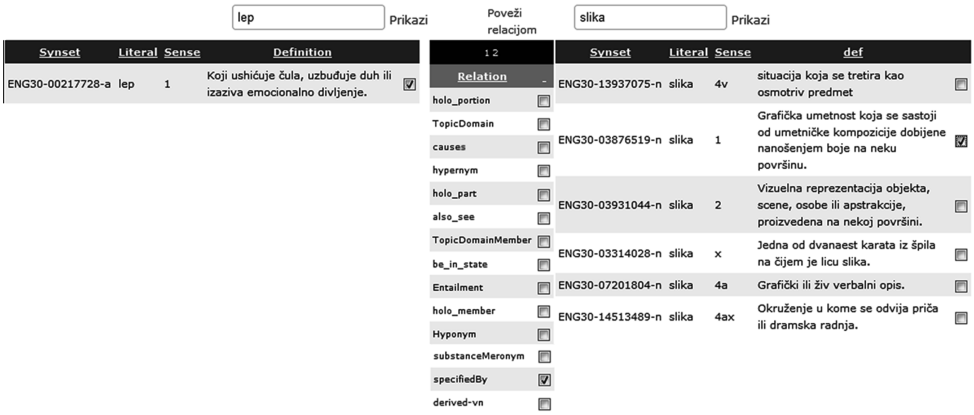


Figure 7: A Web tool for semi-automatic enlargement of SWN; in this case an adjective *lep* (*beautiful*) and a noun *slika* (*picture*) are linked by a pair of inverse relations *specificOf* / *specifiedBy*

"РЕЧ *nonym* ПРИДЕВ ИМЕНИЦА" (WORD *as* ADJECTIVE NOUN). Finally, we built a tool which uses the rule (2) (see Section 3) in the ontology SWN to inference if the extracted linguistic structure represents the figure *simile*.

Linguistic structures extracted from the text collection by regular expressions indicate that comparisons which are uncommon in everyday speech can be found in poetic texts. For example, in the songs of Danojlić, structures like *as good as an elephant*, *calm like a cow*, *shaking like an arrow*, etc. occur. In songs ofRšumović, the following can be found: *blue as the carpet*, *green as carpet*, *ugly as a brush*, etc. At the end of this test, we analyzed linguistic structures extracted from datasets named *news* and *films*, which are used in SAFOS — a system for sentiment analyzing in Serbian language (Mladenović et al., 2016b), to assess the possibility of a further improvement in terms of recognizing the figurative speech. The result showed a low but uniform recognition accuracy[9] of about one third of the total number of appearances of the figure *simile*. In structures that were extracted from these datasets such as: *okarakterisan kao triler* (*characterized as a thriller*), *poznat kao zemlja* (*known as a country*), *urađen kao nastavak* (*made as a sequel*), etc., it can be seen that some of them satisfy the rule (2), but they are not similes and they could be detected as false positive candidates for instancing the figure *simile*. However, the recognition of the

[9] Accuracy is one of the measures of classification quality and represents the ratio of the sum of true positive and true negative classified items and the sum of all items in the test collection.

figures was performed using the knowledge of the ontology SWN, which is the reason why high precision was obtained, and false positive candidates were rejected[10]. On the other hand, unrecognized structures like *oštrim poput brijača* (*sharp as a razor*), *hrabrim kao SUPERMEN* (*brave as Superman*), *pući kao vidik* (*expanded as a view*), *razmnožavati se kao vinska mušica* (*multiplied as Drosophila*) indicate a low level of classification recall and the need for learning different linguistic structures besides those described, taking into account other types of nouns (named-entities, abbreviations, etc.) and parts of speech (verbs, for example). Classification accuracy of figure *simile* was carried out in all texts in a collection made of 10 digitized writings and the two datasets used as testing sets in sentiment analysis in Serbian. The results are shown in Figure 8.

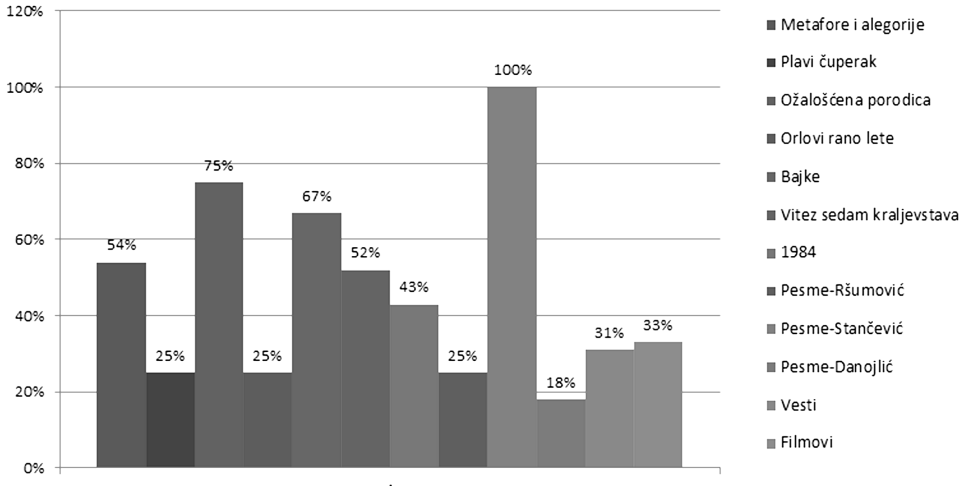


Figure 8: Accuracy of recognition of the figure *simile* in texts in Serbian using the ontology SWN

The evaluation of classification is executed in the process of classifying the figure *simile* in George R.R. Martin's novel "A Knight of the Seven Kingdoms", because that is the largest integral text in the collection. By applying regular expressions we found a total of 135 structures in a form "РЕЧ *као* РЕЧ{1,2}", "РЕЧ *nonym* РЕЧ{1,2}" (WORD *as* WORD{1,2}). From these, a total of 107 structures represented

[10] In the ontology SWN there is not a relation *specifiedBy* between nouns and adjectives, unless it is verified by methods of SWN ontology learning (see Section 4).

the figure simile and 28 did not, like in the examples: "*Tu sam pecao kao dečak*" ("*I caught fish there as a boy*"), "*Lim je imao naoštren kolac koji bi mogao poslužiti kao koplje*" ("*Lem had a sharpened stick that might serve for a spear*"), "*Dve godine je služio kao paž*" ("*He served two years as a page*"), etc. This classification is done by hand, and then a total of 135 structures subjected to ontological classification. This means that for each of 135 pairs (word, noun), a check is performed in order to see if they satisfy the rule (2). Some examples of recognized similes are: "*Bele kao kost bile su put i kosa Brindena Rečnog...*" ("*White as bone were the skin and hair of Brynden Rivers*"), "*Bila su hladna kao kamen, ali ih je bilo divno videti*" ("*They were hard as stone, but beautiful to look upon*"), "*... bilo je crveno kao krv*" ("*was red as blood*"). The results of ontological classifications are shown in Table 1, in the confusion matrix.

figure simile classification		manual	
		yes	no
by classifier	yes	$tp = 42$	$fp = 0$
	no	$fn = 65$	$tn = 28$

Table 1: Confusion matrix that represents classification of linguistic structures extracted by regular expressions from George R.R. Martin's novel "A Knight of the Seven Kingdoms" to those that represent figure *simile* and those that do not.

Based on the confusion matrix, estimates of the system are given in Table 2, and statistical assessment of a classifier by ROC curve is presented in Figure 9, where it is obtained that the $AUC > 0.5$ for a confidence interval of 95%.

Evaluation measures of classification figure *simile*	$Precision = \frac{tp}{tp+fp}$	$Recall = \frac{tp}{tp+fn}$	$F_1 = \frac{2PR}{P+R}$	Accuracy
Measures	1.000	0.393	0.564	0.518

Table 2: Evaluation measures of classification of the figure *simile* in George R.R. Martin's novel "A Knight of the Seven Kingdoms"

Receiver operating characteristic curve (ROC) is one of the methods of evaluating the classifier that uses a graphical representation of the relationship of sensitivity ($\frac{tp}{tp+fn}$) and specificity ($\frac{tn}{tn+fp}$) for each possible score on the test. Figure 9 shows

the (solid line) ROC curve of classification figure simile, while the diagonal line (shown in dashed line, called diagonal accidental outcomes) represents the outcome of a random classification. The area under the ROC curve (AUC) is interpreted as the probability that a randomly selected item labeled as positive (being figure simile), has a higher score than the one labeled as negative. AUC is a measure of the accuracy of the classifier and may have a value of $AUC = 0.5$, when the ROC curve coincides with the diagonal of random outcomes, to $AUC = 1$ in the case of absolute separation of classes. Classification of figure simile in our experiment achieves $AUC = 0.696$.

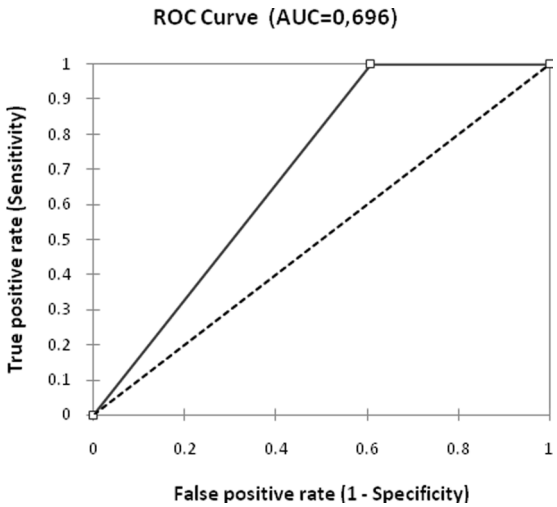


Figure 9: Statistical evaluation of the classification of the figure *simile* in George R.R. Martin's novel "A Knight of the Seven Kingdoms"

The results of the classification show low grade of recall. There are three groups of structures that are not recognized by the proposed system:

1. structures containing personal nouns as given in examples: *pametan kao princ Eris* (*as clever as Prince Aerys*), *visok kao Dank* (*almost of a height with Dunk*), etc;
2. structures that do not contain an adjective and a noun, but a verb and a noun as in examples: "*Onda se okrenu kao vihor, i jurnu u noć*" ("*Whirling, he darted back out into the night*") and "*Jaje ulete kao bez duše*" ("*Egg burst in panting*");

3. structures containing an adjective and a noun, but the order is such that it can not be unambiguously detected by propesed system, and examples of such structures are found in sentences: "*. . . gole kao od majke rođene, brašnjave od glave do pete*", "*. . . čije su vode svetlucale crveno i zlatno, sjajne kao ploča kovanog bakra*", "*Zbog blede kože i kao kreč bele kose ličio je na živi leš*".

Structures containing personal noun can be recognized by including tools for Named Entity Recognition and a pattern in the form of "PRIDEV *kao* VLASTITA_IMENICA" (ADJECTIVE *as* PROPER_NOUN). Those structures containing a verb and a noun can be identified by introducing new paterns like "GLAGOL *kao* IMENICA" (VERB *as* NOUN), and for the third mentioned group of structures, if we would use patterns like "*kao* IMENICA [VEZNIK|PREDLOG]? PRIDEV" (*as a* NOUN [CONJUNCTION| PROPOSAL]? ADJECTIVE).

6 Conclusion

In this paper, we propose a method for recognizing rhetorical figures that belong to the group of figures of twisted meaning (tropes). The method uses rules defined in the ontology based on the semantic network Serbian WordNet to identify whether a linguistic structure extracted from text satisfies some of the rules to be labeled as a rhetorical figure. The method was tested by detecting rhetorical figure simile in a collection of ten digitized writings and the two textual datasets which were previously used in sentiment analysis in Serbian. An evaluation of the system was conducted in the text of the George Martin's novel "Knight of the Seven Kingdoms". Evaluation results ($AUC = 0.696$) showed that the classification of rhetorical figure achieved middle grade, but precision of 100% indicated that further learning of SWN ontology, based on the proposed semi-automatic method, can improve the classification of this figure.

In future work, in the case of classification of the figure *simile*, we will introduce patterns described in section 5 to identify structures which are not recognized by the proposed system. In the case of the other figures, we will work on building of datasets for testing classifiers of rhetorical figures *irony*, *sarcasm* and *oxymoron*. We also plan to use the proposed method in a process of building features in sentiment analysis in Serbian. We will also try to expand the set of ontological rules to include procedures of generalization and specifications of concepts that are already included by the rules laid down by this system.

References

Stefano Baccianella, Andrea Esuli, and Fabrizio Sebastiani. SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining. In Cal-

zolari et al. (2010). ISBN 2-9517408-6-7. URL http://www.lrec-conf.org/proceedings/lrec2010/summaries/769.html.

Krešimir Bagić. *Rječnik stilskih figura*. Školska knjiga, Zagreb, 2012.

Francesco Barbieri, Francesco Ronzano, and Horacio Saggion. UPF-taln: SemEval 2015 Tasks 10 and 11. Sentiment Analysis of Literal and Figurative Language in Twitter. In Daniel M. Cer, David Jurgens, Preslav Nakov, and Torsten Zesch, editors, *Proceedings of the 9th International Workshop on Semantic Evaluation, SemEval@NAACL-HLT 2015, Denver, Colorado, USA, June 4-5, 2015*, pages 704–708. The Association for Computer Linguistics, 2015. ISBN 978-1-941643-40-2. URL http://aclweb.org/anthology/S/S15/S15-2119.pdf.

Daniel G. Bobrow. A Question-Answering System for High School Algebra Word Problems. In *Proceedings of AFIPS Conference, 26. FJCC, Part I*, pages 591–614, New York, 1964. ACM. doi: 10.1145/1464052.1464108.

Nicoletta Calzolari, Khalid Choukri, Bente Maegaard, Joseph Mariani, Jan Odijk, Stelios Piperidis, Mike Rosner, and Daniel Tapias, editors. *Proceedings of the International Conference on Language Resources and Evaluation, LREC 2010, 17-23 May 2010, Valletta, Malta*, 2010. European Language Resources Association. ISBN 2-9517408-6-7. URL http://www.lrec-conf.org/proceedings/lrec2010/index.html.

Paula Carvalho, Luís Sarmento, Mário J. Silva, and Eugénio de Oliveira. Clues for Detecting Irony in User-generated Contents: Oh...!! It's "So Easy" ;-). In *Proceedings of the 1st International CIKM Workshop on Topic-sentiment Analysis for Mass Opinion*, pages 53–56, New York, 2009. ACM.

Paula Carvalho, Luís Sarmento, Jorge Teixeira, and Mário J. Silva. Liars and Saviors in a Sentiment Annotated Corpus of Comments to Political Debates. In Lin et al. (2011), pages 564–568. ISBN 978-1-932432-88-6. URL http://www.aclweb.org/anthology/P11-2099.

Vladan Devedžić. *Semantic Web and Education*. Integrated Series in Information Systems. Springer-Verlag New York Inc., 2010. ISBN 9781441942012.

Richard Farkas, Eszter Simon, Gyorgy Szarvas, and Daniel Varga. Gyder: Maxent Metonymy Resolution. In Eneko Agirre, Lluís Màrquez, and Richard Wicentowski, editors, *SemEval '07: Proceedings of the 4th International Workshop on Semantic Evaluations*, pages 161–164, Stroudsburg, PA, USA, 2007. Association for Computational Linguistics.

Dan Fass. met*: A Method for Discriminating Metonymy and Metaphor by Computer. *Computational Linguistics*, 17(1):49–90, 1991. ISSN 0891-2017.

Christiane Fellbaum. *WordNet: An Electronic Lexical Database*. Language, Speech, and Communication. A Bradford Book, 1998. ISBN 978-0262061971.

Elena Filatova. Irony and Sarcasm: Corpus Generation and Analysis Using Crowdsourcing. In Nicoletta Calzolari (Conference Chair), Khalid Choukri, Thierry

Declerck, Mehmet Uğur Doğan, Bente Maegaard, Joseph Mariani, Jan Odijk, and Stelios Piperidis, editors, *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12)*, Istanbul, Turkey, 2012. European Language Resources Association (ELRA). ISBN 978-2-9517408-7-7.

Jakub Gawryjolek, Chrysanne Di Marco, and Randy A. Harris. An Annotation Tool for Automatically Detecting Rhetorical Figures — System Demonstration. In *Proceedings of the IJCAI-09 workshop on Computational Models of Natural Argument*, Pasadena, CA, 2009.

Roberto I. González-Ibáñez, Smaranda Muresan, and Nina Wacholder. Identifying Sarcasm in Twitter: A Closer Look. In Lin et al. (2011), pages 581–586. ISBN 978-1-932432-88-6. URL `http://www.aclweb.org/anthology/P11-2102`.

Thomas R. Gruber. A Translation Approach to Portable Ontology Specifications. *Knowledge Acquisition*, 5(2):199–220, 1993. ISSN 1042-8143. doi: 10.1006/knac.1993.1008. URL `http://dx.doi.org/10.1006/knac.1993.1008`. Special issue: Current Issues in Knowledge Modeling.

Yanfen Hao and Tony Veale. An Ironic Fist in a Velvet Glove: Creative Mis-Representation in the Construction of Ironic Similes. *Minds and Machines*, 20 (4):635–650, 2010. doi: 10.1007/s11023-010-9211-1. URL `http://dx.doi.org/10.1007/s11023-010-9211-1`.

Andrew Hardie, Veronika Koller, Paul Rayson, and Elena Semino. Exploiting a Semantic Annotation Tool for Metaphor Analysis. In Davies, Matthew and Rayson, Paul and Hunston, Susan and Danielsson, Pernilla, editor, *Proceedings of the Corpus Linguistics 2007 Conference*, Birmingham, 2007. University of Birmingham.

Randy Harris and Chrysanne Di Marco. Constructing a Rhetorical Figuration Ontology. In *Symposium on Persuasive Technology and Digital Behavior Intervention, Convention of the Society for the Study of Artificial Intelligence and Simulation of Behaviour (AISB)*, pages 47–52, Edinburgh, 2009.

Daniel Devatman Hromada. Initial Experiments with Multilingual Extraction of Rhetoric Figures by means of PERL-compatible Regular Expressions. In Irina P. Temnikova, Ivelina Nikolova, and Natalia Konstantinova, editors, *Proceedings of the Student Research Workshop associated with The 8th International Conference on Recent Advances in Natural Language Processing, RANLP 2011, 13 September, 2011, Hissar, Bulgaria*, pages 85–90. RANLP 2011 Organising Committee, 2011. URL `http://www.aclweb.org/anthology/R11-2013`.

Ashley R. Kelly, Nike A. Abbott, Randy Allen Harris, Chrysanne DiMarco, and David R. Cheriton. Toward an Ontology of Rhetorical Figures. In Júnia Coutinho Anacleto, Renata Pontin de Mattos Fortes, and Carlos J. Costa, editors, *Proceedings of the 28th Annual International Conference on Design of Communication, SIGDOC 2010, São Carlos, São Paulo state, Brazil, September 26-29, 2010*, pages 123–130. ACM, 2010. ISBN 978-1-4503-0403-0. doi: 10.1145/1878450.1878471. URL `http://doi.acm.org/10.1145/1878450.1878471`.

Alistair Kennedy and Diana Inkpen. Sentiment Classification of Movie Reviews Using Contextual Valence Shifters. *Computational Intelligence*, 22(2):110–125, 2006. doi: 10.1111/j.1467-8640.2006.00277.x. URL `http://dx.doi.org/10.1111/j.1467-8640.2006.00277.x`.

Svetla Koeva, Cvetana Krstev, and Duško Vitas. Morpho-Semantic Relations in WordNet - a Case Study for two Slavic Languages. In *Proceedings of Global WordNet Conference*, pages 239–253, Szeged, 2008. University of Szeged, Department of Informatics.

Veronika Koller, Andrew Hardie, Paul Rayson, and Elena Semino. Using a Semantic Annotation Tool for the Analysis of Metaphor in Discourse. *Metaphorik.de*, 15: 141–160, 2008.

Johannes Leveling. FUH (FernUniversität in Hagen): Metonymy Recognition Using Different Kinds of Context for a Memory-Based Learner. In Eneko Agirre, Lluís Màrquez, and Richard Wicentowski, editors, *SemEval '07: Proceedings of the 4th International Workshop on Semantic Evaluations*, pages 153–156, Stroudsburg, PA, USA, 2007. Association for Computational Linguistics.

Dekang Lin, Yuji Matsumoto, and Rada Mihalcea, editors. *The 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, Proceedings of the Conference, 19-24 June, 2011, Portland, Oregon, USA - Short Papers*, volume 2, 2011. The Association for Computer Linguistics. ISBN 978-1-932432-88-6.

Katja Markert and Malvina Nissim. Metonymy Resolution As a Classification Task. In *Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing - Volume 10*, EMNLP '02, pages 204–213, Stroudsburg, PA, USA, 2002. Association for Computational Linguistics. doi: 10.3115/1118693.1118720. URL `http://dx.doi.org/10.3115/1118693.1118720`.

Zachary J. Mason. CorMet: A Computational, Corpus-based Conventional Metaphor Extraction System. *Computational Linguistics*, 30(1):23–44, 2004. ISSN 0891-2017. doi: 10.1162/089120104773633376.

Ruslan Mitkov. *Anaphora Resolution*. Longman, Cambridge, UK, 2002.

Jelena Mitrović. Electronic Tools and Resources for Multi-Word Unit Detection and Research in Serbian. The 2th General Meeting of The IC1207 COST Action, PARSEME, Athens, Greece, 10–11 March, 2014, 2014. URL `http://typo.uni-konstanz.de/parseme/images/Meeting/2014-03-11-Athens-meeting/PosterAbstracts/Mitrovic.pdf`.

Jelena Mitrović, Miljana Mladenović, and Cvetana Krstev. Adding MWEs to Serbian Lexical Resources Using Crowdsourcing. The 5th PARSEME General Meeting, Iaşi, Romania, 23–24 September 2015, 2015. URL `http://typo.uni-konstanz.de/parseme/images/Meeting/2015-09-23-Iasi-meeting/WG1-MITROVIC-MLADENOVIC-KRSTEV-poster.pdf`.

Miljana Mladenović and Jelena Mitrović. Ontology of Rhetorical Figures for Serbian. In Ivan Habernal and Václav Matousek, editors, *Text, Speech, and Dialogue - 16th International Conference, TSD 2013, Pilsen, Czech Republic, September 1-5, 2013. Proceedings*, volume 8082 of *Lecture Notes in Computer Science*, pages 386–393. Springer, 2013. ISBN 978-3-642-40584-6. doi: 10.1007/978-3-642-40585-3_49. URL http://dx.doi.org/10.1007/978-3-642-40585-3_49.

Miljana Mladenović and Jelena Mitrović. Semantic Networks for Serbian: New Functionalities of Developing and Maintaining a WordNet Tool. In Gordana Pavlović-Lažetić, Cvetana Krstev, Ivan Obradović, and Duško Vitas, editors, *Natural Language Processing for Serbian: Resources and Applications*, pages 1–11, Belgrade, 2014. University of Belgrade, Faculty of Mathematics. ISBN 978-86-7589-088-1.

Miljana Mladenović, Jelena Mitrović, and Cvetana Krstev. A Language-Independent Model for Introducing a New Semantic Realation between Adjectives and Nouns in a WordNet. In Verginica Barbu Mititelu, Corina Forăscu, Christiane Fellbaum, and Piek Vossen, editors, *The Proceedings of Eighth Global WordNet Conference 2016*, pages 218–225, Bucharest, Romania, 2016a. ISBN 978-606-714-239-6.

Miljana Mladenović, Jelena Mitrović, Cvetana Krstev, and Duško Vitas. Hybrid Sentiment Analysis Framework for a Morphologically Rich Language. *Journal of Intelligent Information Systems*, 46(3):599–620, 2016b. doi: 10.1007/s10844-015-0372-5. URL http://dx.doi.org/10.1007/s10844-015-0372-5.

Cristina Nicolae, Gabriel Nicolae, and Sanda Harabagiu. UTD-HLT-CG: Semantic Architecture for Metonymy Resolution and Classification of Nominal Relations. In Eneko Agirre, Lluís Màrquez, and Richard Wicentowski, editors, *SemEval '07: Proceedings of the 4th International Workshop on Semantic Evaluations*, pages 454–459, Stroudsburg, PA, USA, 2007. Association for Computational Linguistics.

Bo Pang, Lillian Lee, and Shivakumar Vaithyanathan. Thumbs Up?: Sentiment Classification Using Machine Learning Techniques. In *Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing*, volume 10 of *EMNLP '02*, pages 79–86, Stroudsburg, PA, USA, 2002. Association for Computational Linguistics. doi: 10.3115/1118693.1118704. URL http://dx.doi.org/10.3115/1118693.1118704.

Massimo Poesio and Ron Artstein. Anaphoric Annotation in the ARRAU Corpus. In Calzolari et al. (2010), pages 1170–1174. ISBN 2-9517408-6-7. URL http://www.lrec-conf.org/proceedings/lrec2010/index.html.

Vassiliki Rentoumi, Stefanos Petrakis, Manfred Klenner, George A. Vouros, and Vangelis Karkaletsis. United we Stand: Improving Sentiment Analysis by Joining Machine Learning and Rule Based Methods. In Calzolari et al. (2010). ISBN 2-9517408-6-7. URL http://www.lrec-conf.org/proceedings/lrec2010/summaries/41.html.

Antonio Reyes and Paolo Rosso. Building Corpora for Figurative Language Processing: The Case of Irony Detection. In *Proceedings of the 4th International Workshop*

on Corpora for Research on Emotion Sentiment & Social Signals (in conjunction with LREC 2012), pages 94–98, 2012a.

Antonio Reyes and Paolo Rosso. Making Objective Decisions from Subjective Data: Detecting Irony in Customer Reviews. *Decision Support Systems*, 53(4):754–760, 2012b. doi: 10.1016/j.dss.2012.05.027. URL `http://dx.doi.org/10.1016/j.dss.2012.05.027`.

Ekaterina Shutova, Simone Teufel, and Anna Korhonen. Statistical Metaphor Processing. *Computational Linguistics*, 39(2):301–353, 2013. ISSN 0891-2017. doi: 10.1162/COLI_a_00124.

Oren Tsur, Dmitry Davidov, and Ari Rappoport. ICWSM - A great catchy name: Semi-supervised recognition of sarcastic sentences in online product reviews. In William W. Cohen and Samuel Gosling, editors, *Proceedings of the Fourth International Conference on Weblogs and Social Media, ICWSM 2010, Washington, DC, USA, May 23-26, 2010*. The AAAI Press, 2010. URL `http://www.aaai.org/ocs/index.php/ICWSM/ICWSM10/paper/view/1495`.

Miloš Utvić. *Izgradnja referentnog korpusa savremenog srpskog jezika*. Doktorska disertacija, Univerzitet u Beogradu, Filološki fakultet, Beograd, 2014.

Tony Veale. Detecting and Generating Ironic Comparisons: An Application of Creative Information Retrieval. In *Artificial Intelligence of Humor, Papers from the 2012 AAAI Fall Symposium, Arlington, Virginia, USA, November 2-4, 2012*, volume FS-12-02 of *AAAI Technical Report*. AAAI, 2012. URL `http://www.aaai.org/ocs/index.php/FSS/FSS12/paper/view/5557`.

Tony Veale and Yanfen Hao. Support Structures for Linguistic Creativity: A Computational Analysis of Creative Irony in Similes. In *Proceedings of CogSci 2009, the 31st Annual Meeting of the Cognitive Science Society*, pages 1376–1381, 2009.

Lowri Williams, Christian Bannister, Michael Arribas-Ayllon, Alun D. Preece, and Irena Spasić. The Role of Idioms in Sentiment Analysis. *Expert Systems with Applications*, 42(21):7375–7385, 2015. doi: 10.1016/j.eswa.2015.05.039. URL `http://dx.doi.org/10.1016/j.eswa.2015.05.039`.

A Appendix

Collection of 10 digitized writings of various genres (children's literature, novels, fairy tales, comedies and essays) used for testing a proposed method for recognition of rhetorical figure simile.

1. Antić, Miroslav. (1965). Plavi čuperak, accessed 25.02.2016, http://www.kotraza-dragacevo.org/Knjige.html
2. Bajke Za Decu, accessed 25.02.2016, http://www.zadecu.com/Bajke

3. Ćopić, Branko. Orlovi rano lete, accessed 25.02.2016, http://www.kotraza-dragacevo.org/Knjige.html
4. Danojlić, Milovan. (1966). Kako spavaju tramvaji, accessed 25.02.2016, http://borovo1234.weebly.com/uploads/1/2/6/2/12626321/milovan_danojlic_-_kako_spavaju_tramvaji.pdf
5. Martin, George R. R. Vitez sedam kraljevstava. Beograd: Laguna, 2014.
6. Nušić, Branislav. (1935). Ožalošćena porodica, accessed 25.02.2016, https://www.scribd.com/doc/60214229/Branislav-Nusic-Ozaloscena-Porodica
7. Orwell, Gorge. 1984. Beograd: Laguna, 2014.
8. Petrović, Mihailo. Metafore i alegorije. Beograd: Srpska književna zadruga, 1967.
9.Ršumović, Ljubivoje. Ma šta mi reče. Beograd: Laguna, 2013.
10. Stančević, Sreten. „Dečiji osmeh“. Klub pesnika “Orfej”, accessed 25.02.2016, http://www.orfejsu.com/deciji-osmeh-sreten-stancevic-zbirka-decijih-pesama/

Lexical analysis of two-word terminological phrases within distribution system

UDC 811.163.41'373.46
DOI 10.18485/infotheca.2016.16.1_2.3

Tanja Ivanović
tanja.ivanovic09@gmail.com
Government of the Republic of Serbia, European Integration Office

ABSTRACT: This paper presents the analysis of two-word terminological phrases within distribution system, the specifics of word formation as well as the aspects of interrelation between expert terminology and general vocabulary. Certain lexical and semantic principles in power system terminology in Serbian were pointed out through the analysis of paradigmatic relations, as well as the variation i.e. consistency of their phrasal form upon translation into English. Cross-language influences are particulary discussed, especially foreign origin terminology penetration in the Serbian language. Research results indicate that analysed terms used in its syntactic form can hardly be applied outside the terminology circle which they belong to, given its precise reference to a specific phenomenon within the power system, the mode of operation of a specific device or the entire system, an individual part or the wide range of components used in the system, the position or different circumstances peculiar to this profession.

KEYWORDS: two-word terminological phrases, power system, distribution system, the lexeme origin, translation equivalents

PAPER SUBMITTED: 17 March 2016
PAPER ACCEPTED: 04 May 2016

1 Introduction

As a profession with a large rise in the last decade at the global level, the energy has taken a leading role in society which brought about the publishing of a large number of scientific papers and glossaries in this area. Some of the most prominent glossaries in the field of energy are available via the Internet. Such as, for example,

EDF energy – Glossary of terms energy,[1] Bishop Victor-International Glossary of Biochemistry, Construction, Energy & Power Engineering,[2] Glossary of U.S. Energy Information Administration,[3] Glossary of electric industry terms.[4] The development of energy as a discipline is consequently accompanied by a growing number of terminology units. For this reason, a list of terms and their analysis, along with its general importance for business, education and communication, has an essential role in identifying the foreign terms impact on Serbian language (and culture as well) and represents an inseparable component of energy as a scientific discipline.

Although there is a large number of surveys related to general principles in terminology and detailed linguistic analysis of the power system terminology, an incentive for this research was "Linguistic analysis of scientific style of Russian and Serbian language in the field of electrical engineering," a doctoral thesis by Nadežda Lainović-Stojanović (Лаиновић-Стојановић, 1996).

Though frequently used, power distribution terms lack detailed lexical analysis which was the reason and incentive for the research of two-word terminological phrases of power system in Serbian, along with their equivalents in English, in order to bring this lexical system closer to its professional circles, at least to some extent.

This paper was a result of the power distribution utility companies' need for the translation of Electricity Distribution Grid Code into English.[5] As a by-law, this document is mandatory for the standardization of work, planning and maintenance of the power distribution system, respecting the rules of standardized technical terminology in both Serbian and English language.

Given the small number of enlisted terms[6] in the power system in Serbian, the list of these specialized phrases is one of the goals of this paper.

[1] `https://www.edfenergy.com/large-business/glossary`

[2] `http://gen.lib.rus.ec/book/index.php?md5=FD402172420BD18B069D99A7DDB2FBF2`

[3] `http://www.eia.gov/tools/glossary/index.cfm?id=M`

[4] `https://www.xcelenergy.com/staticfiles/xe/Regulatory/EEIGlossaryIRPEEI2005Definitions.pdf`

[5] The Electricity Distribution Grid Code document is available at: `http://www.propisionline.com/IndOk/Legislation/52592`

[6] Apart from the obligation of the European Integration Office of the Republic of Serbia, to translate the entire corpus of the EU *acquis communautaire*, as one of the goals for the Serbia's accession to the European Union, respecting the principle of expert, legal and linguistic review, no other serious attempt of collecting terms in the energy sector has been registered. The only significant attempt in terminology base creation was the manual published within Public Enterprise (PE) Elektroprivreda Srbije in 1997, published by the Electricity Coordination Center. "Statistical terminology used in the electric power industry in Serbian, English and French" was published with a goal to create a unique terminology for the experts in the field of statistics in PE Elektroprivreda Srbije. Although the terminology enlisted in this manual is not comprehensive, it still

2 Lexical analysis

As a part of the vocabulary that includes a set of special words and phrases from a specific scientific, technical and professional field, the terminology is very important for accurate and effective communication within the appropriate specialized language whereas a unified and precise terminology is crucial for proper translation of texts (Влада Републике Србије, Канцеларија за европске интеграције, 2013). As a scientific discipline, the terminology covers three basic concepts such as: basic principles and concepts that underpin the study of terms, the guidelines that are used in terminology work as well as the set of terms of the specific field (Castellví, 1999).

Electricity Distribution Grid Code contains more than a thousand two-word terminological phrases, but only the most frequently used phrases are selected for the purpose of this paper. Appendix at the end of this paper shows terminological phrases in Serbian and English with the origins of its constituents. Note: Columns signifying the origin of the first or the second lexeme are only filled in if lexemes are of foreign origin.

As it was emphasised in this analysis, the most frequent phrases in Serbian are adjective + noun phrases. Phrases such as noun + noun in the genitive case are also often used, and it will be discussed later in the paper.

A few examples of two-word terminological phrases such as adjective + noun, both of foreign origin, are given here:

elektroenergetski sistem	'power system'
energetska analiza	'power analysis'
stacionarni režim	'stationary regime'
energetski transformator	'power transformer'

As a discipline focused on collection, analysis, definition and presentation of terms that belong to a specific professional field, the primary role of terminology is finding the right equivalents for technical terms in the source language i.e. creating relations between terms and concepts that they represent (Bowker, 2015). There is a significant number of international technical terms among the analysed lexemes that retained their meaning from the language they are taken from. Majority of lexemes from the analysed corpus originate from Latin, while a significant number comes from Greek. In a few cases, analysed lexemes originating from Latin or Greek entered the Serbian language through French, German and English. Table 1 shows the lexemes

represents a positive attempt to create a database for some of the most frequently used terms.

of foreign origin with regard to their total number in the analysed corpus, classified according to the language they originate from.

Language	**Lexeme**	**Number of words**	**Perc.**
Latin	activus, apparatus, armarium transformator, coefficiens, consumere, polus, obiectum, condensator, minimalis, maximalis, differentialis, contactus, generator, factor, stationarius, combinare, frequens, directus, magnetizare, distributivus, frequentia, primarius, secundarius	24	11.0%
Greek	analysis, charakteristikos, elektron, energeia, energetikos, kriterion, metron, parallelos, pausis, phasis, synchronos, systema, thermos	13	6.0%
French	batterie, groupe(ital. gruppo), reserve, regime (lat. regimen), electrique (gr. electron)	5	2.0%
German	schema (gr. schema)	1	0.5%
English	impedance (lat. impedire), flick	2	1.0%
total		**45**	**21**%

Table 1: Foreign origin lexemes

In addition to the lexemes shown in Table 1, which are of international origin and which are already incorporated in the Serbian technical language, whilst retaining the meaning from their language of origin, it is important to point out the examples with different meanings of seemingly identical terms which can create confusion in translation because of their "mismatch". Such terms are called "false friends". False friends are words that sound the same but have different meaning in the two languages. Thus, the term regulator which is defined as a device used for automatic voltage control (*naponski regulator* 'voltage controller') in Serbian or device used for the turbine rotational speed control (*turbinski regulator* 'turbine controller'), is marked by the lexeme 'controller' in English.

On the contrary, the lexeme 'regulator' in English defines the regulatory body (institution), such as the Energy Agency. It is a noun of Latin origin and designates "the automatic maintenance of a balanced work" (Anić et al., 2002). Similarly, Serbian equivalent for the term 'capacity' is *instalisana snaga* (rather than *kapacitet*). This term is related to the electric element – capacitor. The term *transformatorska stanica* has an equivalent in English in the concept of 'substation' which, translated as such into Serbian, clearly refers to heat energy and heating rather than power system.

Another example frequently used in language for specific purposes (LSP) is the phrase *upravljanje sistemom* 'system control', where 'control' is not control in terms of supervision or audit but technical management of the system, notwithstanding that it was implemented by various control devices (regulators) or human factor that performs manipulations (connection and disconnection of the system elements) in the system. In this regard, it is noted that there is no unambiguous mapping from one language to another and that the terminology of foreign origin was adopted into Serbian technical language with prior detailed analysis of experts in a given field, relying on the practice in our region for more than a century. Such terminology is, in the analysed corpus of power system profession, now completely adapted and adopted.

The largest number of terminological phrases in analysed corpus of the Serbian language has phrase model adjective + noun which is most frequently translated into English as a phrase model noun + noun, as the most common one in English. Thus, from the above-mentioned examples we can see that only phrase *stacionarni režim* 'stationary regime' is translated into English by the same model adjective + noun, while the rest are translated by noun + noun model.

Moreover, we frequently have cases of foreign origin adjective in a combination with the noun of Slavic origin, such as:

re/aktivna snaga	're/active power'
transformatorsko	polje 'transformer bay'
in/direktno merenje	'in/direct (electricity) metering'
diferencijalna zaštita	'differential protection'
električni luk	'electrical arc'

Though it is not often the case, there are examples of an adjective of domestic and a noun of foreign origin in the analysed corpus.

merna grupa	'metering group'
prenosni sistem	'transmission system'
kružna frekvencija	'angular frequency'

There are several phrases examples in the analysed corpus of Serbian language where both lexemes are of domestic origin:

pad napona	'voltage drop'
vršna snaga	'peak power'

spojno polje	'busbar coupler'
uklopna šema	'topology diagram'
kratak spoj	'short circuit'

Having analysed the phrases, one can conclude that a very small number of examples has synonyms (which can be attributed to the precision and exactness of the analysed discipline), nevertheless, among these examples we have both parallel and equal use of domestic and foreign terms, so that *nazivni*, *naznačeni* or *nominalni napon* in the Serbian language is equivalent to 'rated voltage' in English. Similarly, the term *mala* or *mini hidroelektrana* is translated into English as 'small power plant'. Another example, *jednofazni* or *monofazni priključak* is translated into English as 'single phase connection'. Although the terminology variations and synonyms are typical for standard language, the spontaneous, free and unmotivated use of lexical variants and synonyms is not in the interest of any profession, precisely because it leads to inconsistencies in the use of its terminology (Schmitz and Straub, 2010). Therefore, it is necessary that each area of expertise shall establish its own terminology eliminating ambiguity and variations in terminology. Creation of terminology bases, such as database that contains information about the areas of application of certain concepts and terms that denote them (Melby, 2012), is one of the key elements in the standardization of terminology and basic precondition for achieving high quality in terminology work.[7]

Among terminological phrases of domestic origin, there is a certain number of phrases translated into English concisely, with a phrase consisting of a single lexeme (monolexemic word), such as:

transformatorska stanica	'substation'
razvodno postrojenje	'switchyard'
raskidač (strujnog) kola	'recloser'
instalisana snaga	'capacity'
električna energija	'electricity'
merni (razvodni) orman	'cubicle'

[7] Inter-Active Terminology for Europe (IATE) is the inter-institutional terminology database of the European Union and one of the biggest terminology databases of the EU institutions (`http://iate.europa.eu/SearchByQueryLoad.do?method=load`). Another positive example is Evroteka, the terminology database of the European integration Office of the Government of the Republic of Serbia (`http://prevodjenje.seio.gov.rs/evroteka/index.php?jezik=engl`).

Although there are monolexemic words in Serbian language translated into English as two-word terminological phrases, there are no such examples in the analysed corpus. However, there are multi-word phrases in English with its equivalents in two-word phrases in Serbian, such as:

dalekovodno polje	'overhead line bay'
podzemni vod	'underground power line'
generatorski prekidač	'generator curcuit breaker'
spojni prekidač	'connection circuit breaker'
mala elektrana	'small power plant'
visokoučinski osigurač	'knife blade fuse'
struje zemljospoja	'earth fault current'

As previously mentioned, frequently used phrase model in Serbian noun + noun in the genitive case is translated into English by the phrase model noun + noun, as in the following examples:

pad napona	'voltage drop'
gubitak snage	'power loss'
rastavljač snage	'power disconnector'
predaja energije	'electricity delivery'
mesto priključenja	'connection point'
odvodnik prenapona	'surge arrester'

It is interesting to note that certain examples of the phrase model adjective + noun and noun + noun in the genitive case in Serbian are translated by participle + noun model into English:

koeficijent svođenja	'referring coefficient'
pogonska snaga	'operating power'
rasklopni aparat	'switching device'
merna grupa	'metering group'
grana magnećenja	'magnetizing branch'

A limited number of examples in Serbian have the phrase model adjective (past participle) + noun translated into English as past participle + noun:

odobrena snaga	'approved power'
dozvoljena struja	'allowed current'
kombinovani rad	'combined operation'
naznačena struja	'rated current'
naznačeni napon	'rated voltage'

One fifth of the total number of analysed lexemes are of foreign origin. Only 13% of analysed corpus are two-word phrases where both constituents are of foreign origin, even 44% of phrases is characterized by the lexemes of domestic origin along with the remaining 43% of combined origin (where the first constituent is of foreign and second of domestic origin or vice versa). Most examples represent the phrase model adjective + noun (up to 84%) while the remaining 16% are modeled noun + noun in the genitive case. Figure 1 illustrates the percentage of domestic and foreign origin lexemes presence in the sample of 107 terminological phrases.

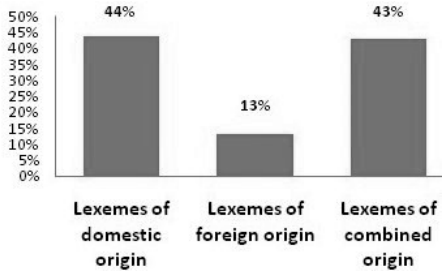


Figure 1. The origin of the phrase constituents in percentage

2.1 Word formation analysis

This section presents the main derivation and compositional abilities of lexemes within the two-word terminological phrases in Serbian and is represented by inventory of affixes within the analysed corpus.

Suffixation Having in mind the suffixation as a morphological process whereby a bound morpheme is attached to the end of a stem in analysed corpus, most examples in Serbian are characterized by adjective suffixes *–an* / *–ni* and *–ski*, while there are only two examples of *–ov* suffix, as in the following examples:

-an/-ni: sinhroni generator, vršna snaga, distributivni objekat/mreža, diferencijalna zaštita, dozvoljena struja, električna energija, električni luk, zaštitni namotaj, zaštitni uređaj, instalisana snaga, karakteristična impedansa, kombinovani rad, kružna frekvencija, konzumno područje, kontaktni termometar, maksimalno/minimalno opterećenje, merna oprema/grupa itd. (51 phrases in total consist of lexemes formed by suffix *-an/-ni*);
-ski: generatorski prekidač, energetski pretvarač/transformator, energetska analiza, kondezatorska baterija, ostrvsko napajanje, ostrvski rad, pogonsko stanje, pogonska snaga, transformatorska stanica/polje;
-ov: Buholcova zaštita, Tevenenova impedansa.

There are a few examples of noun suffixes:

-ač: rastavljač snage, raskidač kola, generatorski prekidač;
-ost: sigurnost napajanja, kriterijum sigurnosti;
-nik: odvodnik prenapona;
-ak: gubitak snage.

Figure 2 illustrates the use of the suffixes in lexemes for 107 analysed phrases.

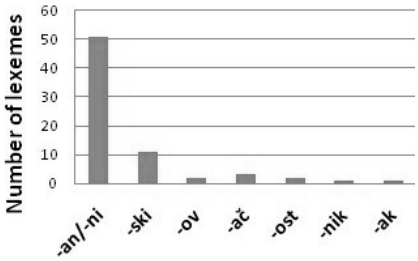


Figure 2. Suffixes

Prefixation Bearing in mind the prefixation as a morphological process of the analysed corpus in Serbian, most frequently used prefixes are of domestic origin: *pod-, nad-, pri-, pre-, raz-* and *bez-*. The emphasis is on the prefixes that form prefix-suffix derivates as shown in the following examples. In the following phrases, the first lexemes are characterized by adjective and prefix both of domestic origin:

beznaponska pauza 'voltage interruption'

nadzemni/podzemni priključak	'overhead/underground connection'
prenaponska	'overvoltage/undervoltage
(nadnaponska)/podnaponska zaštita	protection'
nadpobuđeni/podpobuđeni režim	'underexcitation/overexcitation regime'
prividna snaga	'apparent power'
priključni vod	'connection line'
razvodno postrojenje	'switchyard'
rastavljač snage	'power disconnector'

A combination of domestic origin prefix along with the foreign origin adjective rarely appears.

nadfrekventna/podfrekventna[8] *zaštita*	'overfrequent/underfrequent protection'

The following lexemes from the analysed corpus are formed by prefixes and adjectives, both of foreign origin and, as such, they are adopted in our use:

asinhroni generator	'asynchronous generator'
indirektno merenje	'indirect electricity metering'
reaktivna snaga	'reactive power'

Composition Each compound word is formed by joining stems of two or more separate words into single one, as in the following examples:

zemljospojna zaštita	'earth fault protection'
prekostrujna zaštita	'overcurrent protection'
kratkospojna zaštita	'short-circuit protection'
jednopolna šema	'single-line diagram'
jedno/trofazni priključak	'single/three-phase connection'
visokoučinski osigurač	'knife blade fuse'
poluindirektno merenje	'semi-indirect electricity metering'

[8] Lexemes "podfrekventan" and "nadfrekventan" are taken from the Electricity Distribution Grid Code as an official by-law. As we notice, the voicing assimilation in Serbian was not performed. Therefore, such form is maintained in this paper.

Bearing in mind that some authors considered *polu-* as prefix, primarily because it often serves to partially deny the same stems which prefix *ne-* denies entirely, limiting or mitigating the meaning of the adjective, the categorical affiliation of this formant is hard to determine because this element is very productive and helps in creating new words which is the reason why this formant is analysed as a compound word (Клајн, 2002, pp. 81, 116).[9]

There are several examples of phrases containing compound suffix derivatives:

dalekovodno polje	'overhead line bay'
elektroenergetski objekat	sistem 'power facility / system'

In the analysed corpus of Serbian language, compounds are formed by joining stems by the connecting vowel *-o-*. The only exception was made in the phrase *poluindirektno merenje* where the vowel *-u-* could be considered as a connecting vowel added on the stem *pola* or *po(l)* unless we take into consideration the interpretation of Barić (Barić, 1980, pp. 18–19), where the *polu-* is a prefix (Клајн, 2002, p. 25).

Analysing the total number of lexemes in question, it is easy to perceive the limited number and repetitive character of both prefixes and suffixes that form it (there are nine different forms of prefixes in the analysed corpus along with seven different forms of suffixes), which is the characteristic of the technical field concerned. Although they show greater diversity of formants at first glance, prefixes are much less used for the formation of the analysed corpus (only 17%), while the suffixes are present in 65% of examples. This certainly does not diminish their importance and irreplaceability in the formation of specific terminology database. Furthermore, only eight lexemes of the analysed corpus can be considered the true compounds characterized by clear motivation of both parts of lexemes which is certainly typical for technical field and its exactness, whilst only three phrases contain lexeme which is a compound-derived suffix.

There are 71 lexemes formed by suffix, 17 lexemes formed by prefix and only 11 compound lexemes in the 107 analysed phrases, as illustrated in Figure 3.

[9] An interesting question would be: why this terminology phrase is not set as *poludirektno merenje* instead of the *poluindirektno* because linguistically it is difficult to discern the difference between them. However, there is still a difference because the 'indirect metering' is made through a converter, voltage and current transformers. If we omit one of them, we get 'semi-indirect'. On the other side, as implied by name, the direct metering considers metering without voltage and current transformers via the electricity meter (in this sense, 'semi-direct' as a term in the power system is not used although it would basically mean the same).

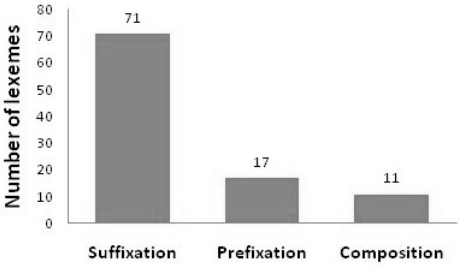


Figure 3. Word formation

3 Lexical and semantic features of terms

Although technical terms analysed by this paper are mostly monosemantic, it is very interesting to analyse their paradigmatic relations too (especially for the polysemantic terms), so the terminology analysis of lexical and semantic approach is addressing one of the most important linguistic issues: issues of polysemy, antonymy and synonymy, which is analysed by this section.

3.1 Polysemic terms

Though the uniformity of terms is of crucial importance in terminology standardization, it is very difficult to implement it completely. Given the fact that the technical terms belong to general vocabulary as well, considering the specificity of the analysed discipline, the right question is whether these terms, i.e. terminological phrases, can in general function independently of the terminology which they belong to?

Certainly, there are examples of terms functioning only in the closed terminology circle. The information as to which terminology field such terms belong to appears in the Dictionary of the Serbian language (DSL), published by Matica Srpska (Матица српска, 2011). For example, the qualifier *fiz.* precedes the lexeme generator which confirms that the term *generator* is the unit belonging to terminology field of physics. In this regard, it is hardly possible to use this term outside its field of terminology. However, due to the fact that the terms are units of the general language, with its defined conceptual content and the new one they develop in contact with other lexical units, when they leave their terminology field and enter the language of journalism and subsequently the standard language, they start to imply its explicit conceptual content, enriching it and developing the relevant elements of implementation (Гортан-Премк, 2004, pp. 122–123). When they develop it (even partly),

they begin with the creation of different lexical systems; they start developing polysemy (Гортан-Премк, 2004, p. 123), such as *generator brojeva* in the lottery, as an example of metonymy.

Terms have limited ability to develop polysemy. Such trait lies in the dualistic nature of the term as a sign of its belonging and the general vocabulary, and thus is subject to all the lexical laws, including polysemy (Гортан-Премк, 2004, p. 119). At the same time, the term belongs to a specific terminology field, having tendency for realization of monosemy as a symmetrical relation between the sign and the concept that is in its semantic content. That is the reason why the terms cannot develop its semantic structure by metaphorical means. However, they can develop it by metonymy and synecdoche very rarely by giving the secondary meaning. In the analysed corpus, we have the examples of the presence of the same term in different terminology fields but with different conceptual values. However, this phenomenon should be distinguished from the homonymy which is based on the similarities of form and disconnectedness of content (Шипка, 2006, p. 61). Actually, homonymy is the existence or the possibility of existence of two or more different semantic contents but according to external appearance, apparently in the same form (Гортан-Премк, 2004, p. 150).

The main reason for polysemy is considered to be the existence of considerably more objects and phenomena in comparison to the number of lexemes that could be used for naming these (Драгићевић, 2010, p. 130). If everything would have a separate title, we would not be able to communicate because we would not be able to remember all the existing lexemes. However, according to this author, the lack of vocabulary should not be characterized by the poverty of one language but by limited cognitive abilities of the language users. The derivation of meaning arises from the basic set of meanings of a lexeme, i.e. its polysemantic structure and consists of primary / basic / designative and secondary meanings (Драгићевић, 2010, p. 131).

Only 37 phrases from the analysed corpus of 107 terminological phrases contain a polysemy noun that is used outside its terminology field. Those are the following: *aparat*, *baterija*,[10] *vôd*,[11] *generator*, *gubitak*, *grana* (*magnećenja*),[12] *grupa*, *energija*,

[10] Noun *baterija* can be considered only in its syntactic form as it appears in the analyzed corpus and that is *kondenzatorska baterija*, because only in this way it undoubtedly refers to the element in the power system, which does not accumulate or convert energy from one form to another (a characteristic of the battery in general is to convert chemical energy into electricity and vice versa), but plays an important role in the compensation of reactive power.

[11] This noun indicates the power line via which the user (power plant or a consumer) is connected to the power system. As such, this noun realizes its secondary meaning.

[12] *Grana magnećenja* is known as lateral or island branch in the power sector. Although it is not present in the Dictionary of Serbian language of Matica srpska Матица српска

zaštita, *kolo*,[13] *lûk*,[14] *mesto*, *merenje*, *mreža*,[15] *napajanje*,[16] *napon*,[17] *objekat*, *oprema*, *opterećenje*, *osigurač*, *pogon*, *područje*, *polje*, *rad*, *režim*,[18] *svođenje*, *sigurnost*, *sistem*, *snaga*, *spoj*,[19] *stanica*, *stanje*, *struja*, *uređaj*, *upravljanje*, *frekvencija*, *šema*.

These 37 polysemy nouns are 17% of the total number of lexemes in the analysed corpus. However, a huge number of polysemic nouns are used more than once in different context, such as: *zaštite* (čak 11), *snage* (takođe 11), *opterećenja* (4), *struja* (5), *vodova* (3), *merenja* (3), *polja* (3), *rada* (3), *režima* (3), *sistema* (3), *generatora* (2), *energije* (2), *mesta* (2), *napajanja* (2), *napona* (2), *objekata* (2), *pogona* (2), *sigurnosti* (2), *stanja* (2), *uređaja* (2) i *šema* (2).

The meaning of the lexemes according to the DSL	**Number of lexeme**
Primary	16
Secondary	21
total	**37**

Table 2: The number of lexemes according to their meaning in DSL

Although analysed lexemes can co-exist in the general vocabulary, if isolated from the phrases we have analysed and used in syntactic relation of LSP, they lose such ability since they undoubtedly point to a particular occurrence in the power

(2011), magnetizing is defined as the process in which electrical energy is transferred by magnetic domains of ferromagnetic core of electric machine or it is defined as core magnetizing (in order to operate properly, the electric machine needs to magnetise its core).

[13] This noun defines the closed conductive flow by which the current circuit is closed: the current circuit.

[14] This noun defines the current flow through the low conductive environment whereby a large amount of heat which has strong light effect is released.

[15] According to DSL, the definition of the term analyzed here is on the 5th place, which means that the lexeme expanded its semantic structure by metaphorical way and got terminological meaning.

[16] This noun does not exist in DSL. It comes from the verb *napajati* which in its primary form represents a continuous verb according to napojiti se, but in its secondary meaning it means to supply, to replenish (the water, electricity etc.) according to DSL.

[17] The difference in the potential between two points is in the current circuit or electric field, which allows the flow of current through the closed circuit.

[18] It refers to a specific system operating condition. In the analyzed corpus there are: *stacionarni režim*, *podpobuđeni režim*, *nadpobuđeni režim*.

[19] In the analyzed corpus the following noun is not enlisted in the DSL: *zemljospoj* (fault current), which defines the current value after one of the three phase conductors touches the zero potential.

system, the position or the wide range of components used in the system. Therefore, the conclusion is that analysed terminology field is in its majority very hermetic. Table 2 shows the number of polysemic lexemes according to its meaning in the DSL.

3.2 Antonimy

While the nouns reflect polysemy in the analysed phrases, the adjectives reflect antonymy as in the examples below:

minimalno opterećenje –	'minimum load' –
maksimalno opterećenje	'maximum (peak) load'
podzemni vod/priključak –	'underground power line/connection' –
nadzemni vod/priključak	'overhead power line/connection'
podpobuđeni režim –	'underexcitation regime' –
nadpobuđeni režim	'overexcitation regime'
podfrekventna zaštita –	'underfrequent protection' –
nadfrekventna zaštita	'overfrequent protection'
prenaponska (nadnaponska) zaštita –	'overvoltage protection' –
podnaponska zaštita	'undervoltage protection'[20]

Whether examples of lexemes given are gradable antonyms or not, they belong to the same lexical and semantic group which means they have the same basic meaning of the word. In most cases, the selected adjectives have complex structure created in word formation, as it is the case with their antonyms.

In addition to that, there are certain examples of phrases presenting lexical antonyms but keeping in mind the difference in power system functioning, they do not have opposing function, so they can be described as complementary antonyms,[21] such as:

jednofazni priključak –	'single phase connection' –
trofazni priključak	'three-phase connection'
primarna struja –	'primary current' –
sekundarna struja	'secondary current'
primarni namotaj –	'primary winding' –
sekundarni namotaj	'secondary winding'

[20] There are two types of voltage protection and accordingly, the antonymy relationship cannot be: voltage – undervoltage or voltage – overvoltage protection.

[21] Reflecting the similarity in diversity of non-linguistic reality as well as the vocabulary part that it reflects, the antonymy embodies the opposite between the two lexemes. In accordance with the nature of this contradiction, it is possible to distinguish at least five basic types of antonyms - gradable, complementary, diametric, reciprocal and reversive (Prćić, 1997, p. 102).

direktno merenje –	'direct electricity metering' –
indirektno merenje –	'indirect electricity metering' –
poluindirektno merenje	'semi-indirect electricity metering'
aktivna snaga –	'active power' –
reaktivna snaga	'reactive power'
sinhroni generator –	'synchronous generator' –
asinhroni generator	'asynchronous generator'

In its prototypical manifestation, complementary antonymy expresses the impossibility of comparing the observed situations and things (Prćić, 1997, p. 105).

Among the analysed phrases, there is an example: *dozvoljena struja* that actually implies the 'maximum allowed current' where the maximum is understood and thus often excluded from the phrase, even in written texts. It also does not have its own antonym, which means that there is no 'inadmissible current' or 'minimum allowed current'. The same goes for *odobrena snaga* where its antonym in the general vocabulary is 'unapproved', and does not exist as such in the analysed field. In analysed corpus there is also an example: *prekostrujne zaštite*, while, analogous to the above examples, 'undercurrent protection' does not exist. The same holds true for *visokoučinski osigurač* which is translated as 'knife blade fuse' and has no antonym as *niskoučinski osigurač*.

Antonymy appears in basic lexemes; precisely in those lexemes demonstrate polysemy and derivation and which exclude synonymy (as a secondary lexical phenomenon). This leads to a conclusion that antonymy, together with variation (polysemantic and derivation), is the main mechanism in the organization of the lexical system (Гортан-Премк, 2004, p. 149). The same holds true for the lexical system of power sector as well.

In relation to the total number of examples in the corpus of 107 analysed terminological phrases, the 12 antonym pairs are selected (out of a total of 25 terminology phrases), which means that the antonymy occurs in 23% of cases in relation to the total number of phrases. It means that antonymy is largely the foundation for the formation of this terminology circle. The examples of synonymy in electricity or distribution terminology are rare, perhaps due to the specific nature of this discipline that is inclined to the greatest possible accuracy and precision. Only two examples can be selected in the analysed corpus where both lexemes are of domestic origin:

prenaponska zaštita – nadnaponska zaštita	'overvoltage protection'
grana magnećenja – poprečna grana – otočna grana	'magnetizing branch'

Both domestic and foreign origin terms are used as equivalents in translation into Serbian for the phrases: 'rated voltage', 'small power plant', 'single phase connection', as synonyms, which has already been discussed.

Table 3 shows an overview of the analysed lexemes according to their formative abilities for the sample of 107 analysed phrases.

4 Cross-language impacts and penetration of terminology from other languages

Even though modern experts in the field of power sector believe that every technical term in foreign language needs to have an appropriate equivalent in Serbian, there are a number of terms that are international and, as such, have been already incorporated in the vocabulary of modern Serbian language and adapted to its script and structure. This, of course, has numerous advantages such as the unification of terms which facilitates communication among scientists and experts in lectures, conferences and business contacts and also facilitates the use of technical and scientific literature. Taking into consideration the fact that language is a living process, this phenomenon is quite natural and expected, especially for the modern time, which is, among other things, characterized by the global connection of distant peoples and languages. Moreover, one of the key reasons why some terms are taken over is the dominance of the language of the nations whose science has a leading role in a global society.

Word formation mode			**No. of lex.**	**Perc.**
Compounds			8	7.5%
Compound suffix derivatives			3	3.0%
Prefix (prefix-suffix) derivatives: *nad-*, *pod-*, *raz-*, *pri-*, *in-*, *bez-*, *a-*, *in-*, *re-*			18	17.0%
Suffix word formation	Noun suffixes		7	6.5%
	Adjective Suffixes:	*-an/-ni*	49	46.0%
		-ski	11	10.0%
		-ov	2	1.9%
Polysemantic lexemes			37	34.5%
Antonymy			25	23.0%
Synonymy			5	4.7%

Table 3. Terminology phrases shown by the number of lexemes and their percentage in the analysed corpus

As regards Serbian language, it was strongly influenced by English in the past few decades, and the presence of certain terms which we can say are uncritically adopted has been noted. However, the attempts of modern power engineers to find the right equivalent in Serbian for the technical terms in English are encouraging. One of the most prominent examples of uncritically adopted term is 'recloser' which was at first translated as an *inteligentni linijski prekidač* and then as *uklopnik* (keeping its original form). Due to the fact that energy as a professional field is rich with devices that interrupt electrical circuits which are referred to as circuit breakers, it is clear that *inteligentni linijski prekidač* was not clearly pointing to the object it was representing and it was therefore necessary to make a difference in relation to the other switches. Thus, the term *raskidač* was accepted as its equivalent in Serbian language.

5 Conclusion

Although every scientific field can boast of distinct terminology, there are not many technical fields experiencing such expansion in recent years as it is the case with the power sector. Along with its development, but adapting to the trends and needs of modern society, there is a new, specific vocabulary which enriches our standard language at the same time through different media.

Considering the specific nature of this technical and scientific sector which is characterized by a universal tendency towards more precise and more exact, this feature is found on the morphological level where each phrase, used in the singular (or plural) is translated by the same number into English. Deviations from the general accuracy are found in the examples of two-word phrases in Serbian, which are translated into English by a single lexeme or in the examples of multi-word structures as their English equivalents are the two-word phrases in Serbian language.

Analysing the examples of the corpus, it is concluded that phrase constituents (standard lexemes) can coexist in the general lexical system even isolated from the analysed phrases, independently, yet they would lose such ability if used in syntactic relation of LSP because, since as such, they lack an inherent language generally and unambiguously signify a specific occurrence within the power system, the position or the devices necessary for the functioning of the system. This is supported by the fact that such words are used with the noun and function as attributes thus forming noun phrases characteristic for the power sector. For this reason, it is very difficult to use these terminological phrases independently of the terminology field which they belong to and thus they remain closed in its terminology circles. These phrases represent the unity of the concept and performance, making internal "form" of language (Белић, 1958), and we can freely add LSP.

In relation to the total number of analysed lexemes of the phrases in this paper, only a fifth of the entire collection of terms are of foreign origin (mostly taken directly from Latin and Greek, or taken over from French, German and English and adapted into Serbian). Only 13% of the analysed phrases consist of foreign origin lexemes, even 44% is characterized by the phrases in which both lexemes are of domestic origin while 43% is of combined origin, in the analysed corpus of 107 terminological phrases. Although the number of prefixes or suffixes is reduced to only a few different formants (there are nine different forms of prefixes and seven different forms of suffix in the analysed corpus), prefixes are less used for the formation of the analysed terminology corpus (17%) while the suffixes are present in even 65% of examples, compared to the total number of analysed lexemes. Nevertheless, although their derivation role is not the same in the formation of specific terminology base, it is equally important. In addition, only 11 lexemes of this selection have complex structure which is characterized by clear motivation of both lexemes in a phrase. A limited number of lexemes in the phrases analysed (only 37 of them), demonstrates polisemy. The characteristic of closed terminology circle is even confirmed by the examples of synonymy that are very rare (only five lexemes showed this trait) in the analysed corpus, while the antonymy is one of the most important mechanisms in the organization of the lexical system of the analysed professional field (23% of the analysed corpus).

Since analysed terminological phrases reflect a comprehensive knowledge of the power system field that provide a linguistic and conceptual accuracy by its semantic characteristics, even though their lexemes show a tendency to create antonyms, rarely creating synonyms and resisting polysemy, this research has shown that their exclusive reference to its own terminology circle is undubitable and that different relationships cannot be established in neither standard nor even in the informal use.

Given that this study is done on a small sample of terminological phrases, this linguistic analysis provides only a partial response to lexical issues analysed in this paper. Although modest in contribution, this paper represents an important starting point and a good foundation for further translation of official documents in power sector such as technical recommendations and internal standards that shall supplement the Electricity distribution Grid Code. In addition, it represents a solid incentive to consider lexical issues in professional circles in order to reach precise determination of the meaning of certain terms and terminology standardization within the specialized profession.

References

Anić, Vladimir, Dunja Brozović-Rončević, Ivo Goldstein, Slavko Goldstein, Ljiljana Jojić, Ranko Matasović, and Ivo Pranjković. *Hrvatski enciklopedijski rječnik.* Za-

greb : Novi liber, 2002.

Barić, Eugenija. *Imeničke složenice i neprefiksalne i nesufiksalne tvorbe.* Zagreb : Liber, 1980.

Белић, Александар. *О језичкој природи и језичком развитку: лингвистичка испитивања.* Београд : Нолит, 1958.

Bowker, Lynne. "Terminology and translation". In: *Handbook of Terminology*, edited by Hendrik J. Kockaert and Frieda Steurs, Vol. 1, pp 304–323, Amsterdam: John Benjamins Publishing Company, 2015.

Cabré, M Teresa. *Terminology: Theory, methods and applications.* Amsterdam : John Benjamins Publishing, 1999.

Драгићевић, Рајна. *Лексикологија српског језика.* Београд : Завод за уџбенике и наставна средства, 2010.

Гортан-Премк, Даринка. *Полисемија и организација лексичког система у српскоме језику.* Београд : Завод за уџбенике и наставна средства, 2004.

Клајн, Иван. *Творба речи у савременом српском језику, први део: слагање и префиксација.* Београд : Завод за уџбенике и наставна средства, 2002.

Лаиновић-Стојановић, Надежда. „Лингвистичка анализа научног стила руског и српског језика у области електротехнике (лексички ниво)". Докт. дисертација, Универзитет у Београду, 1996.

Матица српска. *Речник српског језика.* Нови Сад : Матица српска, 2011.

Melby, Alan K. "Terminology in the age of multilingual corpora". *The Journal of Specialised Translation*, 18(2012):7–29. http://www.jostrans.org/issue18/art_melby.php.

Prćić, Tvrtko. *Semantika i pragmatika reči.* Sremski Karlovci; Novi Sad : Izdavačka knjižarnica Zorana Stojanovića, 1997.

Schmitz, Klaus-Dirk and Daniela Straub. *Successful terminology management in companies: practical tips and guidelines: basic principles, implementation, cost-benefit analysis and system overview.* Stuttgart : Tekom, 2010.

Шипка, Данко. *Основи лексикологије и сродних дисциплина.* Нови Сад : Матица српска, 2006.

Влада Републике Србије, Канцеларија за европске интеграције. *Приручник за превођење правних аката Европске уније.* Београд: Република Србија, Влада, Канцеларија за европске интеграције, 2013.

Appendix – Terminological phrases in Serbian and English with words origins[22]

	Terminological phrase in Serbian	The origin of the first lexeme	The origin of the second lexeme	Terminological phrase in English
1	aktivna snaga	lat. *activus*		active power
2	asinhroni generator	gr. *synchronos*	lat. *generator*	asynchronous generator
3	beznaponska pauza		gr. *pausis*	voltage interruption
4	Buholcova zaštita			Buchholz protection
5	visokoučinski osigurač			knife blade fuse
6	vršna snaga			peak power
7	vršno opterećenje			peak load
8	generatorski prekidač	lat. *generator*		generator circuit breaker
9	grana magnećenja		nlat. *magnetizare*	magnetizing branch
10	gubitak snage			power loss
11	dalekovodno polje			overhead line bay
12	direktno merenje	lat. *directus*		direct electricity metering
13	distributivna mreža	nlat. *distributivus*		distribution network
14	distributivni objekat	nlat. *distributivus*	lat. obiectum	distribution facility
15	diferencijalna zaštita	nlat. *differentialis*		differential protection
16	dozvoljena struja			(maximum) allowed current
17	električna energija	fr. *electrique*	gr. *energeia*	electricity
18	električni luk	fr. *electrique* from grč. *elektron*		electrical arc
19	elektroenergetski objekat	gr. *elektron*; grč. *energetikos*	lat. *obiectum*	power facility
20	elektroenergetski sistem	gr. *elektron*; grč. *energetikos*	gr. *systema*	power system
21	elektroenergetska analiza	gr. *energetikos*	gr. *analysis*	power analysis
22	energetski pretvarač	gr. *energetikos*		power converter
23	energetski transformator	gr. *energetikos*	nlat. *transformator*	power transformer
24	zaštitni namotaj			protective winding
25	zaštitni uređaj			protective device

[22] Columns signifying the origin of the first or the second lexeme are only filled in if lexemes are of foreign origin

	Terminological phrase in Serbian	The origin of the first lexeme	The origin of the second lexeme	Terminological phrase in English
26	zemljospojna zaštita			earth fault protection
27	indirektno merenje	lat. *indirectus*		indirect electricity metering
28	instalisana snaga	srlat. *installatio*		capacity
29	jednopolna šema	lat. *polus*	germ. *schema*	single-line diagram
30	jednofazni priključak monofazni priključak	gr. *phasis*		single phase connection
31	karakteristična impedansa	gr. *charakteristikos*	engl. impedance from lat. *impedire*	characteristic impedance
32	koeficijent svođenja	nlat. *coefficiens*		referring coefficient
33	koeficijent flikera	nlat. *coefficiens*	engl. *flick*	flicker coefficient
34	kombinovani rad	lat. *combinare*		combined operation
35	kondezatorska baterija	nlat. *condensator*	fr. *batterie*	capacitor battery
36	konzumno područje	lat. *consumere*		consumption area
37	kontaktni termometar	lat. *contactus*	gr. *thermos* gr. *metron*	contact thermometer
38	kratak spoj			short circuit
39	kratkospojna zaštita			short-circuit protection
40	kriterijum sigurnosti	gr. *kriterion*		security criterion
41	kružna frekvencija		lat. *frequentia*	angular frequency
42	maksimalno opterećenje	nlat. *maximalis*		maximum / peak load
43	mala elektrana mini elektrana		fr. *electrique*	small power plant
44	merna grupa		fr. *groupe* from ital. *gruppo*	metering group
45	merna oprema			metering equipment
46	merni (razvodni) orman		srlat. *armarium*	cubicle
47	merni namotaj			metering winding
48	merni uređaj			metering device
49	merno mesto			metering point
50	mesto priključenja			connection point
51	minimalno opterećenje	nlat. *minimalis*		minimum load
52	nadzemni vod			overhead line
53	nadzemni priključak			overhead connection

	Terminological phrase in Serbian	The origin of the first lexeme	The origin of the second lexeme	Terminological phrase in English
54	nadpobuđeni režim		fr. *regime* from lat. *regimen*	overexcitation regime
55.	nadfrekventna zaštita	lat. *frequens*		overfrequent protection
56	nazivni napon naznačeni napon nominalni napon			rated voltage
57	naznačena struja			rated current
58	naponska zaštita			voltage protection
59	odvodnik prenapona			surge arrester
60	odobrena snaga			approved power
61	ostrvski rad			island operation
62	ostrvsko napajanje			island supply
63	pad napona			voltage drop
64	paralelni rad	gr. *parallelos*		parallel operation
65	pogonska snaga			operating power
66	pogonsko stanje			operating state
67	podzemni vod			underground line
68	podzemni priključak			underground connection
69	podnaponska zaštita			undervoltage protection
70	podpobuđeni režim		fr. *regime* from lat. *regimen*	underexcitation regime
71	podfrekventna zaštita	lat. *frequens*		underfrequent protection
72	poluindirektno merenje	lat. *directus*		semi-indirect electricity metering
73	poremećeni pogon			operation disturbance
74	pouzdan pogon			reliable operation
75	predaja (električne) energije		gr. *energetikos*	electricity delivery
76	prekostrujna zaštita			overcurrent protection
77	prenaponska zaštita nadnaponska zaštita			overvoltage protection
78	prenosni sistem		gr. *systema*	transmission system
79	prividna snaga			apparent power
80	priključni vod			connection line
81	primarna struja	lat. *primarius*		primary current
82	primarni namotaj	lat. *primarius*		primary winding

	Terminological phrase in Serbian	The origin of the first lexeme	The origin of the second lexeme	Terminological phrase in English
83	razvodno postrojenje			switchyard
84	raskidač (strujnog) kola			recloser
85	rasklopni aparat		lat. *apparatus*	switching device
86	rastavljač snage			power disconnector
87	reaktivna snaga	lat. *reactivus*		reactive power
88	rezervna zaštita	fr. *reserve*		reserve protection
89	sekundarna struja	lat. *secundarius*		secondary current
90	sekundarni namotaj	lat. *secundarius*		secondary winding
91	sigurnost napajanja			security of supply
92	sinhroni generator	gr. *synchronos*	lat. generator	synchronous generator
93	snaga transformatora		nlat. *transformator*	transformer power
94	spojni prekidač			connection circuit breaker
95	spojno polje			usbar coupler
96	stacionarni režim	klat. *stationarius*	fr. *regime* from lat. *regimen*	stationary regime
97	struje zemljospoja			earth fault current
98	strujno opterećenje			current load
99	Tevenenova impedansa		engl. *impedance* from lat. *impedire*	Thevenen's impedance
100	transformatorska stanica	nlat. *transformator*		substation
101	trafo (transformatorsko) polje	nlat. *transformator*		transformer bay
102	trofazni priključak	gr. *phasis*		three-phase connection
103	uklopna šema		germ. *schema*	topology diagram
104	uklopno stanje			topology condition
105	upravljanje sistemom		gr. *systema*	system control
106	faktor poremećaja	lat. *factor*		disturbance factor
107	faktor snage	lat. *factor*		power factor

The Bibliometric and Citation Analyses of the *SPATIUM* Journal

UDC 001.3:050 SPATIUM "2009/2015"
DOI 10.18485/infotheca.2016.16.1_2.4

Milena Milinković
milenam@iaus.ac.rs
Institute of Architecture and Urban & Spatial Planning of Serbia

ABSTRACT: The paper gives the bibliometric and citation analyses of the *SPATIUM* journal for the period 2009-2015. In the subject period, *Spatium* was published twice or three times a year, from the issues 19-34, so that this study comprises 16 volumes, or 145 papers. The bibliometric analysis of the Journal was made *de visu* (with the publication in hand), thus ensuring the authenticity of the given results. The analysis shows the number of papers, categorization of papers, number of authors, affiliation of authors, representation of authors from different countries, average number of references per paper, length of texts, keyword frequency, total number of reviewers and their international representation. The citation count was based on the data taken from the citation databases of SCIndex (Serbian Citation Index), Google scholar and Scopus. The paper shows the differences in the citation analysis of the *SPATIUM* journal at the example of one paper.

KEYWORDS: scientific journal, Spatium, bibliometric analysis, citation analysis, Serbia

PAPER SUBMITTED: 24 February 2016
PAPER ACCEPTED: 28 March 2016

1 Introduction

One of the main conditions for continuity and advances in each field of science in the contemporary conditions includes a reliable, high-quality and regularly published scientific journal. Such journal provides timely information about the latest achievements in a specific scientific field.

And, while the printed publications (first the books, then also journals) were dominant sources of scientific information, and the research results were recorded on paper, with the development of modern technologies since the end of the 20th century, the electronic publications have been increasingly represented, both those published in electronic form as a separate physical entity (floppy disc, CD, DVD etc.) and those available on the Internet.[1]

In the contemporary world, the scientific information is not considered relevant unless published. Hence, the scientific journals in either print or electronic format will have the role of universal communication channels and mediators in the relevant scientific community for many years. In order for a paper to be published in a scientific journal, it should be positively assessed by reviewers, which implies that it must meet the previously set criteria related to the originality and quality of the text. The reviewers from the relevant scientific fields that correspond to the theme of the paper are chosen by the journal's editorial board and based on the scientific reputation the reviewers enjoy in the scientific community.

A scientific journal is a medium for preserving the scientific information and it has an important role in the formation of scientific authority, presentation of the research results and evaluation of professional and scientific contributions and scientific research quality. The papers published in scientific journals are the most important scientific communication channels. Hence, the role of scientific articles is not only to convey scientific information about the definite and unchangeable knowledge, but also to encourage the scientists by their contents to exchange ideas, as well as to encourage them to further research (Вучковић, 2009).

2 *SPATIUM* Journal

The *SPATIUM* journal[2] has been published by the Institute of Architecture and Urban & Spatial Planning of Serbia (hereinafter referred as "the Institute") over 19 years. In the period from 1997 to 2015, 34 issues were published within 31 volumes, containing 261 papers and 23 contributions (book overviews, conference overviews,

[1] The Law on the Obligatory Copy of Publications defines an electronic publication as "a publication published in electronic form as a separate physical entity (diskette, CD, DVD, etc.), a publication available on the Internet and a publication prepared for printing in a format which is in accordance with international standards of universal availability of information" (Закон о обавезном примерку публикација („Службени гласник РС", бр. 52/2011), accessed 15.02.2015, `http://bds.rs/dokumenti/Zakon%20obavezni%20primerak%202011.pdf`).

[2] Spatium / editor in chief Miodrag Vujošević. (Belgrade: Institute of Architecture and Urban & Spatial Planning of Serbia, 1997-), br. 1–34 (1997–2015)

obituaries, information about symposiums, translations of previously published papers, etc.). Out of the total number of volumes, 28 were single volumes, while the remaining 3 were double volumes (Table 1).

Year	Issue	Year	Issue
1997	1, 2	1998	3, 4
1999	5	2000	6
2001	7	2002	8
2003	9	2004	10, 11
2005	12	2006	13/14
2007	15/16	2008	17/18
2009	19, 20, 21	2010	22, 23
2011	24, 25, 26	2012	27, 28
2013	29, 30	2014	31, 32
2015	33, 34		

Table 1. Periodicity of Publication

Until 2009, all papers got only one positive review each and were not classified in any of the categories. Since 2009, when the *Act on Scientific Journal Editing*[3] entered into force, the criteria for journal editing have become significantly stricter. The Editorial Board has begun to strictly implement the *Rules of Procedure and the Manner of Evaluation and Quantitative Presentation of Scientific-research Results of Researchers*[4] (hereinafter referred to as "The Rules") and since then, the criteria for publication of scientific papers include the positive opinions of two reviewers for scientific papers and one positive review for professional papers. The mutual anonymity of authors and reviewers is taken into account, while reviews are performed by the most eminent experts in the specific scientific fields, both from the country and abroad.

[3] The Ministry of Education, Science and Technological Development adopted the Act on Scientific Journal Editing (Record No.: 110-00-17/2009-01, of 09.07.2009) with the aim to assist editorial boards in improving the quality and national scientific periodicals, thus ensuring greater inclusion of journals into the Scientific Information System both at the national and international levels. Accessed 19.01.2016, `http://kobson.nb.rs/upload/documents/MNTR/Dokumenti/akt_o_uredjivanju_casopisa.pdf`

[4] Правилник о поступку и начину вредновања, и квантитативном исказивању научноистраживачких резултата истраживача („Сл. гласник РС", бр. 38/2008)

At the time when the Journal was started in 1997, it was conceived as a scientific organ primarily for the scientific workers of the Institute and their associates. Given that the cooperation with the colleagues from abroad was hampered during the period of international isolation of Serbia, the starting of such scientific journal should have enabled the overcoming of this problem and the establishment of an easier communication and exchange of the scientific and professional thought.

The papers in the Journal are represented across different fields: spatial planning, urban planning, architecture, landscape architecture, geodesy, sustainable development and environmental protection, housing, urban renewal, urban development, issues related to the public utilities and housing, cultural and natural heritage, capital construction, information systems, strategic management, etc.

The *SPATIUM* journal is published in English language and classified into the M24 category (national journal of international importance) according to the classification of national scientific journals into the category of Transport, Urban Planning and Civil Engineering. The confirmation of the quality and international importance of the Journal lies in the fact that in October 2010 the Ministry of Education, Science and Technological Development of the Republic of Serbia included the Journal in the DOI[5] (Digital Object Identifier) system that enables the identification of documents in electronic format and creation of a persistent link to the location of the original document on the Internet. The DOI names have been assigned to all papers since 2002. Precisely thanks to DOI names, the papers, their authors and journals in which they are published have become „visible“ to the professionals and scientific community, which significantly contributes to the international affirmation of authors and the Journal itself.

In 2012, the Institute was under contractual obligations with *Versita* as co-publisher of *SPATIUM*. *Versita* is actually an electronic database on the European scientific and profession journal publishers (Central European Science Publishers)[6] offering an online technology for surveying the contents of journals, paper summaries, cited references, papers in the full-text format, assistance in the evaluation and global promotion of journals, which should result in a significantly greater visibil-

[5] A digital object identifier (DOI) is a serial code used to uniquely identify electronic documents. The DOI for a document is permanent and thus more stable method for referring to an online document than the URL.

[6] “Since its foundation, the Versita company, owned by Mr. Jacek Ciesielski, with registered seat in Warsaw, has strived to be the leading commercial Central European publisher of scientific journals. It has strived to achieve this by launching new Central European journals, on the one hand, and by converting already renowned Central European journals to online versions provided that journals are published in English language, in which they have succeeded to a great extent, primarily in Poland, and then also in Slovakia, Serbia, Czech Republic and Croatia” (Tóth, 2007)

ity, greater readership and increased citation impact. Only in 2013, *Versita* published 15,000 open access articles. From the beginning of 2014, De Gruyter is a new owner of *Versita* and the company changed its name into *De Gruyter Open*[7] with a plan to additionally extend its activities outside Europe, to the USA and Asia, in the forthcoming period.

The ultimate goal of the Editorial Board is to enter the SPATIUM journal in the Thomson Reuters List. Each year Thomson Reuters assesses approximately 2,000 new journals on which it sends its report to the Web of Science[8] (WoS), and the pass rate is up to 10–12 percent. The evaluation is free of charge and the evaluation criteria are strictly defined. After applying for evaluation, it is necessary to send the next three volumes immediately after their publication. After three consecutive issues have been received, the editors of corresponding database evaluate the publications.

The papers published in the *SPATIUM* journal are available in the full-text format through the following databases, catalogues and services:

– DOI Serbia[9] – for the papers published in the period 2002–2015.[10]

[7] `http://www.degruyter.com/view/j/spat`. Taken on 02.09.2014

[8] "Web of Science®, to which the Serbian academic community is subscribed, contains three main citation indexes: Science Citation Index Expanded (SCIE) – the database in the fields of natural, biomedical and technical sciences; Social Sciences Citation Index (SSCI) – the database encompassing the journals in the fields of social sciences; Arts & Humanities Citation Index (AHCI) – the database of journals in the fields of art and humanistic sciences. Since October 2008, it also contains the conference proceedings databases: Conference Proceedings Citation Index Science (CPCIS) since 2001, the citation index for the natural and technical sciences; Conference Proceedings Citation Index- Social Science & Humanities (CPCI-SSH), the citation index for the social and humanistic sciences since 2001. The Journal Citation Report (JCR) is a specific database within the Web of Knowledge℠ platform. It was created by processing the results from the mentioned citation indexes. The JCR contains data on journals classified into thematic categories, within which they are ranked according to their impact factors (IF)" (Антоњћ et al., 2009)

[9] "The DOI Serbia is a collection of scientific journals published in Serbia. The full-text digital archive encompasses the period from 2002 onwards. The journals are processed from cover to cover. All reported journals are published by professional associations, and some of them are also indexed in WoS and CA. The megadata, as well as the full texts, are processed according to the OAI-PMH, and the megadata download is free." Taken on 21.01.2016 from `http://www.digitallibrary.eu/tel4/collection/a0155?locale=sr`

[10] Taken on 01.02.2016 from `http://www.doiserbia.nb.rs/journal.aspx?issn=1450-569X`

– Academic Journals Database[11] – for the papers published in the period 2002–2013.
– Serbian Citation Index[12] - for the papers published in the period 2000–2013.
– Google Scholar[13] (mostly citied papers are available) – for the papers published in the period 2002–2015.
– DOAJ[14] – Directory of Open Access Journals – for the papers published in the period 2002–2015.

2.1 Editors

The twelve-year period of its publication (1997–2015) was marked by the engagement of two guest editors. The double issue 17/18 and issue number 19 had guest editors in 2008 and in the first half of 2009. The position of deputy editor-

[11] The survey of contents, summaries, as well as full text papers, published in the SPATIUM journal is available on the website of the Academic Journals Database, the catalogue of scientific publications in different fields. Taken on 01.02.2016 from `http://journaldatabase.info/journal/issn1450-569X`

[12] "The SCI is a Serbian national citation index developed to serve as a supplement to the international citation indexes. It indexes national journals categorized as periodical scientific publications. The SCI currently contains 1,009,058 references from 67,657 articles, out of which 30,924 in the full text, which have been published in 199 national journals since 2000, in the field of humanistic sciences since 1996, while in the field of social sciences since 1991 onwards." Taken on 17.03.2016 from `http://scindeks.ceon.rs/`

[13] "Compared to the WOS and Scopus, the Google Scholar is a database freely accessible to anyone online. This database encompasses the data on the contents of journals and other publications that publishers placed on their sites, as well as the data from digital repositories, personal web pages, blogs of prominent professionals, preprints, etc. The citations are automatically extracted from the open access texts. The greatest advantage of this database lies in the fact that it is available for free to everyone, while the greatest disadvantage regarding the bibliometric investigation lies in the fact that the database scope is not known. The Google does not provide information about where from it collects the citation data. The errors also occur given that the data are collected automatically. If there is a full-text paper in several places online and in several versions, it also appears in this database so that it is not clear to which of the versions the quoted citations refer to." (Raičević, 2013)

[14] "The Directory of Open Access Journals is a service that provides access to the quality controlled Open Access Journals, namely to the free electronic journals that have met specific scientific and academic criteria. The articles of a great number of journals are available in full text. The Directory covers a plenty of scientific fields, while the ultimate goal is that all scientific disciplines, in all important languages, are represented". Taken on 21.01.2016 from `https://doaj.org/toc/1450-569X`

in-chief was introduced in 2009 from the issue number 20, the same issue when the Editor-in-Chief was relieved (Table 2).

Editors	Year	Journal issue
Nada Milašin, Editor-in-Chief	1997	1, 2
	1998	3, 4
	1999	5
	2000	6
	2001	7
	2002	8
	2003	9
	2004	10, 11
	2005	12
	2006	13/14
	2007	15/16
Nada Milašin, Editor-in-Chief;	2008	17/18
Miodrag Vujošević, **Jasna Petrić** - Guest editors	2009	19
Miodrag Vujošević, Editor-in-Chief; **Jasna Petrić**, Deputy Editor-in-Chief	2009	20, 21
	2010	22, 23
	2011	24, 25, 26
	2012	27, 28
	2013	29, 30
	2014	31, 32
Miodrag Vujošević, Editor-in-Chief; **Tamara Maričić**, Deputy Editor-in-Chief	2015	33, 34

Table 2. Journal Editors

With the change of Editorial Board members and, later, the entry into force of the *Act on Scientific Journal Editing*, the criteria for journal editing were significantly tightened. From March 2009 (since the issue number 19), all professional and scientific papers have contained abstracts, keywords, introduction, illustrations and tables that follow the text, and the references.

The articles are regularly classified into appropriate categories and reviewed by competent professionals with scientific knowledge.

The mentioned change of Editorial Board members, tightened editing criteria and strict implementation of the *Act on Scientific Journal Editing* present a turning point in the editing policy, this being the reason why this analysis covers exactly the period of time spanning from 2009 to 2015.

3 The Sample and Methodological Approach

As already mentioned, the sample for bibliometric analysis comprises all papers published in the period from 2009 to 2015. In this seven-year period, the journal issues 19–34 were published and they contained the total of 145 papers. The data necessary for this investigation were collected using the inductive method, while the bibliometric analysis of the Journal was performed *de visu* (with the publication in hand), thus ensuring the authenticity of the quoted data.

The citation analysis was done based on the citation data taken from the citation databases of Google Scholar, SCIndex and Scopus. The citation count for each individual volume was taken from the Google Scholar citation database and from the citation database of the SCIndex for the period 2009–2015, while from the Scopus citation database for the period from 2011, when the *SPATIUM* indexing in this database started, to 2015. In order to more easily notice the difference in the number of citations in dependence on the database, as well as to also indicate a need for consulting all available sources in the evaluation and ranking of the journal, but also in the evaluation of author impact factor, this paper shows the differences in the citation analysis of the *SPATIUM* scientific journal at an example of the paper entitled *Modelling the spatial distribution of Vojvodina's population by using dasymetric method* by Nikola Krunić, Branislav Bajat, Milan Kilibarda and Dragutin Tošić, published in 2011 in the issue number 24.

4 Bibliometric Analysis

The models, types of analyses and tables used for the bibliometric analysis of *SPATIUM* journal were taken from the literature that deals with the evaluation of scientific journals and bibliometric investigations in different fields of science. The paper shows the results related to the presented number of papers (Martek and Šute, 2010; Tella and Aisha Olabooye, 2014), categorization of papers (Martek and Šute, 2010), number of authors (Thanuskodi, 2010; Hussain et al., 2011; Jena et al., 2012), affiliation of authors (Thanuskodi, 2010; Hussain et al., 2011; Jokić and Zauder, 2013), representation of authors from different countries (Jena et al., 2012), average number of references per paper, text lengths (Thanuskodi, 2010; Tella and Aisha Olabooye, 2014) and the total number of peer-reviewers and their international representation.

4.1 Number of Papers per Volume

The total of 145 papers was published in the covered period. The number of published papers per journal issue ranges from 7 to 13. Seven papers where published four times, in issues 19, 23, 26 and 27, and the largest number of papers within a volume (13) were published in issues 31 and 33 (Table 3).

Year	Journal issue	Number of published papers per volume	Total number of papers per year
2009	19	7	25
	20	10	
	21	8	
2010	22	8	15
	23	8	
2011	24	10	25
	25	8	
	26	7	
2012	27	7	17
	28	10	
2013	29	12	22
	30	10	
2014	31	13	20
	32	7	
2015	33	13	21
	34	8	
Total		**145**	**145**

Table 3. Number of papers per volume

4.2 Categorization of Papers

Out of 145 papers that were published in the period 2009–2015, all the papers were reviewed, 144 articles were classified into the category of scientific and professional papers, while one of the papers was not categorized. According to *The Rules*, the articles published in journals are classified into the following categories:

Category of scientific papers

1. Original scientific paper;

2. Review scientific paper;
3. Short or preliminary communication;
4. Scientific critique or polemics.

Professional papers

1. Scientific paper;
2. Informative contribution;
3. Survey papers.

Exceptionally, the papers in certain areas of science that are published in journals can be monographic studies or critical editions.

Out of the total number of papers, 116 are in the category of scientific papers, 27 in the category of professional papers, 1 paper is classified as "Monographic study" and 1 article is not categorized but published as a technical report. The survey papers (69) account for the largest number of papers in the category of scientific papers, followed by the original scientific papers (30), while 9 papers are classified into the category of short or preliminary communication. Finally, 8 papers are classified into the category of scientific critique or polemics. Three contributions were published in issues 26, 27 and 32. The first contribution contains an information about the TURaS international scientific project, the second one contains an information about the conference RESPAG that was organized by the Institute of Architecture and Urban & Spatial Planning of Serbia, while the third contribution is a translation of the paper entitled *Social space as the subject of scientific research – Spatium*, by Milorad Macura, published in 1965 in the journal *Savremene urbanističke teme*, Issue 2, by the Institute of Architecture and Urban & Spatial Planning of Serbia (pp. 43–63), (Table 4).

4.3 Lengths of Papers

Figure 1 shows the representation of papers of different lengths in each of the analysed issues of the *SPATIUM* journal herein. The largest number of published papers, 110 (75.86%), are papers published on 5–8 pages, less than one fourth of the total number of papers, 26 (17.93%) were published on 9–12 pages, while the number of texts published on less than 4, or more than 13 pages is negligible.

4.4 Authorship Analysis

The authorship analysis of papers published in the period 2009–2014 shows that there were 48 single-authored papers and 97 co-authored papers (Table 5) out of which 45 were two-authored papers, 38 three-authored papers, 9 four-authored papers and 8 papers were written by five- and more authors (Table 6). Regarding the

SPATIUM	Categorized papers						NP*	Contributions**
	OSP	RSP	SPC	SCP	MS	PP	TR	
19 (2009)	4	2			1			
20 (2009)	1	5	1	2		1		
21 (2009)	3	5						
22 (2010)		5	1	1		1		
23 (2010)	1	5	1					
24 (2011)	1	5	2			2		
25 (2011)	1	4		1		2		
26 (2011)	1	3				3		1
27 (2012)		3	1			3		1
28 (2012)	4	6						
29 (2013)	2	4	1	4		1		
30 (2013)	4	3	1			2		
31 (2014)	5	5				2	1	
32 (2014)	1	3				3		1
33 (2015)	2	3	1			7		
34 (2015)		8						
Total	**30**	**69**	**9**	**8**	**1**	**27**	**1**	**3**
Percentage	20.69	47.59	6.21	5.52	0.68	18.62	0.68	
Total	145 papers (100%)							3

Table 4. Categorized and non-categorized papers

*NP – Noncategorized papers;
**Contributions – Contributions (information on the projects and conferences, translations of previously published papers);
OSP – Original scientific paper; RSP – Review scientific paper; SPC – Short or preliminary communication; SCP – Scientific critique or polemics; MS – Monographic study; PP – Professional papers; TR – Technical report

multi-authored papers, the two- and three-authored papers prevail, while the number of four- and five-authored papers is negligible. What can be concluded from the authorship analysis is that there were no significant oscillations in the number of single-authored papers, neither were there significant oscillations in the number of co-authored papers in the period 2010–2014, as well as that the relationship between them is approximately the same, while it may be observed that there is a difference as regards the papers published in 2009 and in the last three journal issues (32, 33 and 34). In the issues number 19, 20 and 21 published in 2009, the number of co-authored papers was four times greater relative to the number of single-authored papers, while

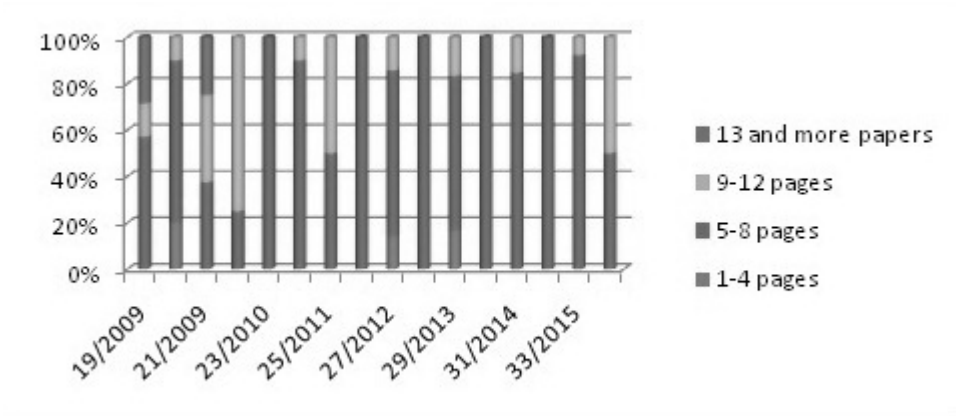


Figure 1. Representation of papers of different lengths

in issues 32, 33 and 34, out of 28 papers, only three were single-authored papers. The number of authors per paper is important because, in ranking authors for the election into the academic ranks, the maximum points are awarded to theoretical papers written by maximum three co-authors, to numerical simulations by five co-authors and to experimental papers by seven co-authors (Ковачевић, 2009).

Authorship	**Year**							**Total**	**Percentage**
	2009	2010	2011	2012	2013	2014	2015		
Single author	5	6	12	6	10	8	1	**48**	**33.90**
Co-authorship	20	9	13	11	12	12	20	**97**	**66.90**
Total	25	15	25	17	22	20	21	**145**	**100.00**

Table 5. Relationship between single- and co-authored papers

4.5 Institutional Affiliation of Authors

The investigation into the affiliation of authors was carried out for three institutions: Institute of Architecture and Urban & Spatial Planning of Serbia (IAUS), which is a publisher of the journal, the Faculty of Architecture (FA), University of

SPATIUM	**Number of authors per paper**					**Total**
	1 author	2 authors	3 authors	4 authors	5 and more authors	
19 (2009)	3		4			
20 (2009)	1	4	3	2		
21 (2009)	1	2	5			
22 (2010)	3	2	3			
23 (2010)	31	4				
24 (2011)	3	6		1		
25 (2011)	6	1			1	
26 (2011)	3	3	1			
27 (2012)	3	3	1			
28 (2012)	3	4	3			
29 (2013)	6	3	3			
30 (2013)	4	2	2	1	1	
31 (2014)	6	1	4	1	1	
32 (2014)	2	2	2	1		
33 (2015)	1	5	4	1	2	
34 (2015)		3	3	2		
Total	**48**	**45**	**38**	**9**	**5**	**145**
Percentage	**33.10**	**31.03**	**26.21**	**6.21**	**3.45**	**100.00**

Table 6. Number of single- and co-authored papers per volume

Belgrade, and the Faculty of Geography (FG), University of Belgrade, with its Department of Spatial Planning, which were interesting for this study because of the nature of the Institute's activities, while the number of authors from other institutions in Serbia and from abroad are given in Table 7 in the aggregate.

The total of 320 authors, out of which 226 authors from Serbia and 94 authors from abroad, published their papers in the 16 analysed volumes. In the subject period, 66 authors were from the Institute, 52 authors were from the Faculty of Architecture, while only 2 authors were from the Faculty of Geography. The remaining 106 authors from Serbia were either employed in some of the institutions (99) or individuals associated with some of the institutions (7).

SPATIUM	IAUS	FG	FA	Other institutions in Serbia	Institutions from abroad	Individuals (from Serbia)	Total no. of authors from different institutions
19 (2009)	7	1	4	1	2		15
20 (2009)	1		9	12	4		26
21 (2009)	6			6	8		20
22 (2010)	7		6	1	2		16
23 (2010)	2		2	1	6		11
24 (2011)	4	1	5	3	4	2	19
25 (2011)			5	9	2		16
26 (2011)	5		3	1	2	1	12
27 (2012)			2	4	3	3	12
28 (2012)	9		3	6	2		20
29 (2013)			3	4	14		21
30 (2013)	2		1	12	10		25
31 (2014)	7		4	10	8	1	30
32 (2014)	4		1	10	1		16
33 (2015)	7		1	7	23		38
34 (2015)	5		3	12	3		23
Total	**66**	**2**	**52**	**99**	**94**	**7**	**320**
Percentage	**20.63**	**0.63**	**16.25**	**30.94**	**29.38**	**2.19**	**100.00**

Table 7. Institutional affiliation of authors

4.6 Number of National and International Papers

As shown in Figure 2, there are 96 (66.21%) papers written by national authors, while the remaining 49 (33.79%) papers by international authors[15]. This confirms the fact that rather many international papers were published in the journal.

Further investigation indicates that papers were sent by authors from different countries of the world, out of which the most represented authors were from Europe, but there were also authors from Africa, Asia and North and South America. The papers written by authors from Serbia were the most numerous (96), followed by papers written by authors from Greece (13) and papers by authors from Slovenia, Great Britain and Germany (4). What is important for the journal itself, its popularity and reputation is the fact that the number of authors from abroad is growing.

4.7 References Overview

The list of cited and used literature given at the end of the text is one of the main conditions for the categorization of each paper. Such list is important for the quality

[15] All papers by internation authors were taken into account , as well as co-authored papers by national and international authors.

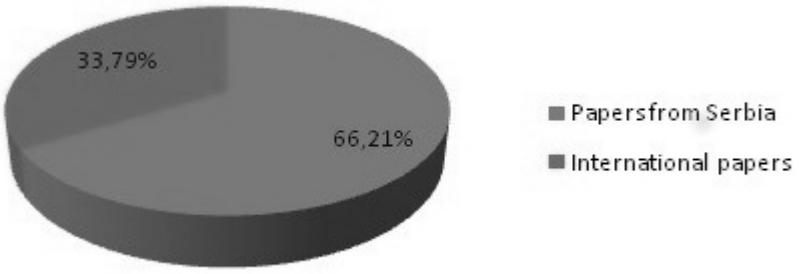


Figure 2. Number of national and international papers

of an article because it provides an insight into the material an author used during his/her research and which directly instigated and encouraged him/her to undertake a specific research, as well as which affected the course of his/her research. But the list is also equally important for those scientific workers who will use the research in future. By tracing the references in historical sequence, from the used literature, the literature referred to and so on backward, it is possible to comprehend the field covered by the paper and become familiar with the specific research (Đurđević, 2010).

The most cited scientific papers have the greatest impact on the development of a particular science. The greatest number of the most cited authors of such papers are the authors who have published important results obtained from their research, and which their colleagues recognize as something original and stimulating, as something that stands out for its scientific importance and contribution.

Table 8 depicts the number of the categorized papers with references per volume and the total number of the categorized papers with references in the period from 2009 to 2015, the total number of references per journal issue and total number of references for all papers published in the seven-year period. Finally, it depicts the average number of literature referred to in each volume and average number of references for all papers in the *SPATIUM* journal in the covered period of time.

The published 144 professional and scientific papers have 4,158 references, which makes an average of 28.88 references per published paper. This analysis did not include the average age of references, neither did it include kind of referred literature or exclude the self-citation by authors. This opens a possibility for further investigation, but also indicates the fact that authors used literature to a large extent, which is an important precondition for a serious and comprehensive research work.

SPATIUM	Year																Total
	2009			2010		2011			2012		2013		2014		2015		
	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	
A*	7	10	8	8	7	10	8	7	7	10	12	10	12	7	13	8	**144**
B*	113	163	311	220	142	274	313	191	220	314	408	293	323	197	352	324	**4158**
C*	16.14	16.3	38.88	27.5	20.29	27.4	39.13	27.29	31.43	31.4	34	29.3	26.92	28.14	27.08	40.5	**28.88**

Table 8. Number of references of the categorized papers

A* Total number of categorized papers with list of references
B* Total number of references
C* Average number of references per paper

Figure 3 shows the relationship between the number of papers and the average number of references for each volume of the analysed journal, in the period 2009–2015. It can be observed that the pattern of reduction or increase in the number of references relative to the number of papers published in issues 21–33 is almost identical and, with some oscillations, this number is around the average number of references counted for the entire period (28.88). More significant deviations are observed in issues 19 and 20 when the average number of the pieces of literature referred to was smaller (16.14 and 16.3 respectively) and in the last analysed volume (34) in which the number of the pieces of literature referred to significant increased to an average of 40.5 references per paper.

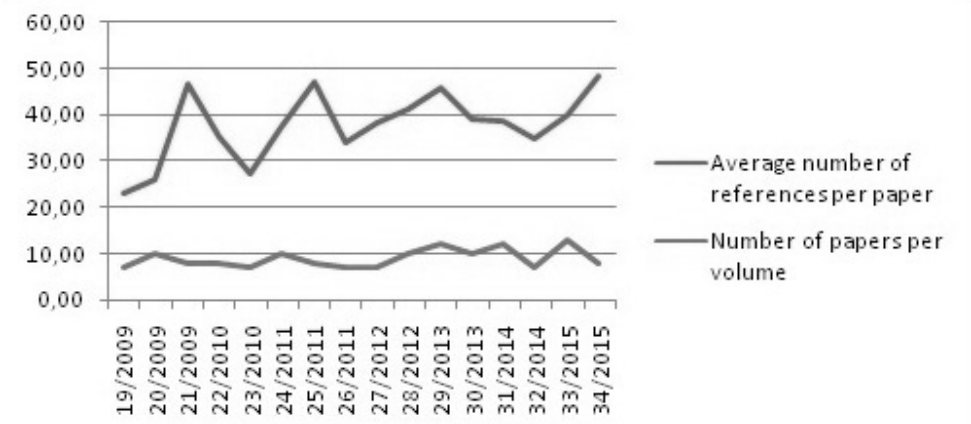


Figure 3. Relationship between the number of papers and the average number of references per volume

4.8 Keyword Analysis

Defining the right keywords is of importance for a summarized presentation of contents of a scientific paper (Jevremov, 2009). They are a basis for the classification of documents and their indexing in different databases.

The sample used for the keyword analysis comprised the papers published in the period 2009–2015. Out of 145 published papers, 3 papers did not have the keywords (one paper was published in 2009 in issue 20, and one was published in issue 22 and one in issue 23, both in 2010), which is a high-level representation of keywords in the subject body (Table 9). It was found that there were 673 keywords in the remaining 142 papers, out of which 627 keywords were used only once, while the number of repeated keywords in the remaining 46 differed: 28 keywords were repeated twice, 9 keywords were repeated three times, 3 keywords were repeated four times and 4 keywords were repeated five times, while out of 2 keywords, 1 was repeated eight times and 1 was repeated ten times. The two most repeated keywords were: *Serbia* (10) and *sustainable development* (8). *The environment*, *identity*, *spatial planning* and *sustainability* were four terms that were used five times as keywords. The following keywords were repeated three times: *Belgrade*, *climate change* and *planning* (Figure 4).

The authors determined the keywords in their papers by themselves and, given that there is no a controlled glossary in the domain of spatial planning, the non-standardized use of terms was observed.

Year	Number of keywords	Number of papers with keywords	Number of papers without keywords
2009	125	24	1 (20)
2010	66	13	1 (22) 1 (23)
2011	117	25	
2012	73	17	
2013	104	22	
2014	90	20	
2015	98	21	
Total	673	142	3

Table 9. Keywords

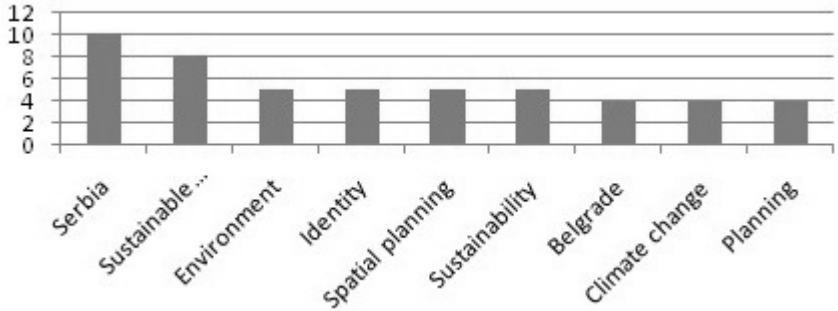


Figure 4. The most frequently repeated keywords

4.9 Reviewers

Prior to publishing, the quality and validity of the presented scientific paper must be confirmed by an expert who is competent in the specific field of paper. The Editorial Board of *SPATIUM* takes care that reviews are entrusted to the most eminent experts from the country and abroad, also taking into account the mutual anonymity of both the authors and the reviewers. The constant advocacy of Editor-in-Chief and Deputy Editor-in-Chief for attracting as many reviewers as possible contributes to a higher status of the Journal. Table 10 depicts the number of reviewers from Serbia and abroad per each of the sixteen issues covered by this investigation, in the period from 2009 to 2015.

SPATIUM	**Year**																**Total**	**Percentage**
	2009			2010		2011			2012		2013		2014		2015			
	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34		
Reviewers for Serbia	7	11	11	11	8	17	9	11	8	15	17	15	23	14	18	7	**202**	**71.13**
Reviewers from abroad	0	6	1	4	5	2	6	1	5	6	3	4	6	4	14	15	**82**	**28.87**
Total	7	17	12	15	13	19	15	12	13	21	20	19	29	18	32	22	**284**	**100**

Table 10. Reviewers

5 Citation Analysis

The citation counts for the published papers in which final results of certain phases of scientific research are presented is one of the main criteria in evaluating the scientific manuscripts. The citation analysis can be used for evaluating an individual researcher, universities and scientific institutions as a whole, as well as for evaluating the journals in which scientific papers are published (Raičević, 2013).

This type of evaluation can be applied to each field of science. The creation of a citation index has resulted from a desire and need to create a database that would enable keeping the track of scientific research, scientific ideas and advancement and development of a particular science. Each scientific worker uses literature relevant for his/her research. Citing the papers, the scientists link the research results to the previously published results (Filipi Matutinović, 2013).

The promotion of researchers and their election into the higher academic ranks is largely dependent on the number of times their papers were cited. Despite the shortcomings in the evaluation of scientific papers based on the citation data, such as great number of self-citations[16] and co-citations, citation manipulation within the scientific circles, lack of an adequate link between the specific paper and the cited papers, motives to cite or not to cite that do not have to be, as a rule, the scientific ones, etc., there are also positive sides because frequent citation of an author is a recognition that the author's paper had a great impact on subsequent research. Furthermore, the citations are publicly available and confidential, and the citation collection and citation analysis are greatly facilitated by the citation databases.

5.1 Data Sources for the Citation of Papers Published in the *SPATIUM* Journal

The data on the citation of papers published in the SPATIUM journal were taken from the following citation databases:

– SCIndex (Serbian Citation Index) – the citation database and search engine for accessing scientific journals published in Serbia;

[16] In the citation count, the relationship between citations and self-citations is important. The data that the number of self-citations is greater than the usual 20 percent relative to the total number of citations can indicate an attempt of manipulation. "The high self-citation rate is typical for the leading scientific journals both because of permanently high quality of papers published in them and because of the unique or thematic proliferation... However, there are journals that have a high self-citation rate so that they deform the overall perception of citation of such journals...in case of significant deviations...ISI checks whether it is a matter of self-citation which is done for the purpose of raising the impact factor of the journal." (Vukasović, 2009)

– Google Scholar – the largest citation search engine;
– Scopus – owned by Elsevier international publishing company.

The time span within which the citation data for the papers published in the *SPATIUM* journal are available:

– Serbian Citation Index – for papers published in the period from 2000 to 2013;
– Google Scholar – for papers published in the period from 2000 to 2015;
– Scopus[17] – for papers published in the period from 2011 to 2015.

Table 11 gives the citation data on the *SPATIUM* journal in the period from 2009 to 2015, for issues 19–34. The total number of citations in the SCIndex and Google scholar citation databases is given first for the period from 2009 to 2015, and then, in parentheses, for the period from 2011 to 2015, which corresponds to the total number of citations in the Scopus citation database, for the purpose of allowing the comparisons between these three citation sources. It can be observed that each of the citation sources contains a different number of citations for a specific journal issue in a certain year, and thereby the number of citations of a particular paper in an issue is different. So it is important to perform the citation analysis for a scientist from Serbia on the basis of all available sources and to never make a comparison between the data obtained for this scientist from a particular database and the data obtained from other database for some other scientists (Filipi Matutinović, 2013).

SPATIUM	**Year**																**Total**
	2009			2010		2011			2012		2013		2014		2015		
	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	
Total number of papers	7	10	8	8	7	10	8	7	7	10	12	10	12+1	7	13	8	**145**
SCIndex	12	8	23	10	14	19	8	8	3	1	1	0	-	-	-	-	**107 (40)**
Scopus	-	-	-	-	-	13	13	9	8	8	4	9	4	5	0	0	**73 (73)**
Google Scholar	17	31	62	34	36	32	18	15	16	12	10	9	4	2	1	-	**299 (119)**

Table 11. Citation count

[17] This database provides the possibility to authors to see the citation counts for their own papers using the Citation Tracker. It is very important that there is the possibility to exclude self-citations using the option Exclude Self-Citations. The citation data for a particular article are given in the Excel table showing the number of citations for a particular article for each year, starting from 1996, as well as the total number of citations.

5.2 Differences in the Number of Citations at the Example of One Paper

At the example of the paper entitled *Modelling the spatial distribution of Vojvodina's population by using dasymetric method*, by co-authors Nikola Krunić, Branislav Bajat, Milan Kilibarda and Dragutin Tošić, we will show the difference in the number of times the paper was cited and where and when it was cited based on the data taken from three different sources. The paper was published in 2011 in the issue number 24. The data were taken on the same day (19 January, 2016) from three citation databases: SCIndex, Google Scholar and Scopus.

SCIndex Citation Overview As it can be seen, the paper was cited three times in the SCIndex database. In 2011, it was cited in the *Collection of Papers of the Faculty of Geography, University of Belgrade*. In 2012, the paper was cited in the *SPATIUM* journal and in 2014 in the *Geonauka* journal. All three papers were also shown as results in the Google Scholar citation database, while out of the above mentioned papers, the paper entitled *Spatial-functional organization of settlements in Vojvodina*, 2012, was shown as a result in the Scopus citation database. In all three cases, the paper was cited by national authors, out of which one is a citation and two are self-citations (Figure 5)[18].

Google Scholar Citation Overview The paper was cited 9 times in the Google Scholar citation database. The papers are neither listed in historical sequence nor in an alphabetical order, but in order in which the search engine has pulled the data from the world networks. The papers were published in the period from 2011 to 2015. According to the Google Scholar citation database, the paper *Modelling the spatial distribution of Vojvodina's population by using dasymetric method* was cited twice in 2011, and once in the Collection of Papers and once in the Journal in 2012. In 2015, it was cited once in a doctoral dissertation and once in a chapter of a monographic publication. All three papers are found amongst these nine papers and they are also shown as results in the SCIndex citation database. Out of the total number of citations, the paper *Modelling the spatial distribution of Vojvodina's population by using dasymetric method* was cited in the Google Scholar citation database, 7 times by national authors 7 and twice by international authors, out of which 6 were self-citations and 3 were citations (Figure 6)[19].

[18] `http://scindeks.ceon.rs/Related.aspx?artcit=1450-569X1124045K`

[19] `http://scholar.google.com/scholar?cites=12858484065582622O5&as_sdt=2005&sciodt=0,5&hl=sr`

Figure 5. SCIndex Citation Overview

Scopus's Citation Overview By getting an insight into the results obtained from the Scopus citation database, we can see that the paper entitled *Modelling the spatial distribution of Vojvodina's population by using dasymetric method* was cited four times in the Scopus citation database, one time each year of the period from 2012 to 2015. In all three cases, this paper was cited in the papers written by national authors and all three were self-citations, while in the fourth case, the paper was cited in the paper by an international author (Figure 7)[20].

Further analysis shows that the paper entitled *Spatial-functional organization of settlements in Vojvodina*, 2012, was cited in three citation databases. The paper entitled *Dasymetric modelling of population dynamic in urban areas*, 2013, was obtained as the result only from the Scopus citation database. The papers entitled *A fine-scale spatial population distribution on the High-resolution Gridded Population Surface and application in Alachua County, Florida* and the *Dasymetric Mapping of Population Distribution in Serbia Based on Soil Sealing Degrees Layer* were obtained as results in the Scopus citation database and in the Google Scholar search engine.

[20] `http://www.scopus.com/results/citedbyresults.url?sort=plf-f&cite=2-s2.0-84857317689&src=s&imp=t&sid=0B17A6BC49186401A0CD58721DF71637.CnvicAmOODVwpVrjSeqQ%3a280&sot=cite&sdt=a&sl=0&origin=resultslist&txGid=0B17A6BC49186401A0CD58721DF71637.CnvicAmOODVwpVrjSeqQ%3a28`

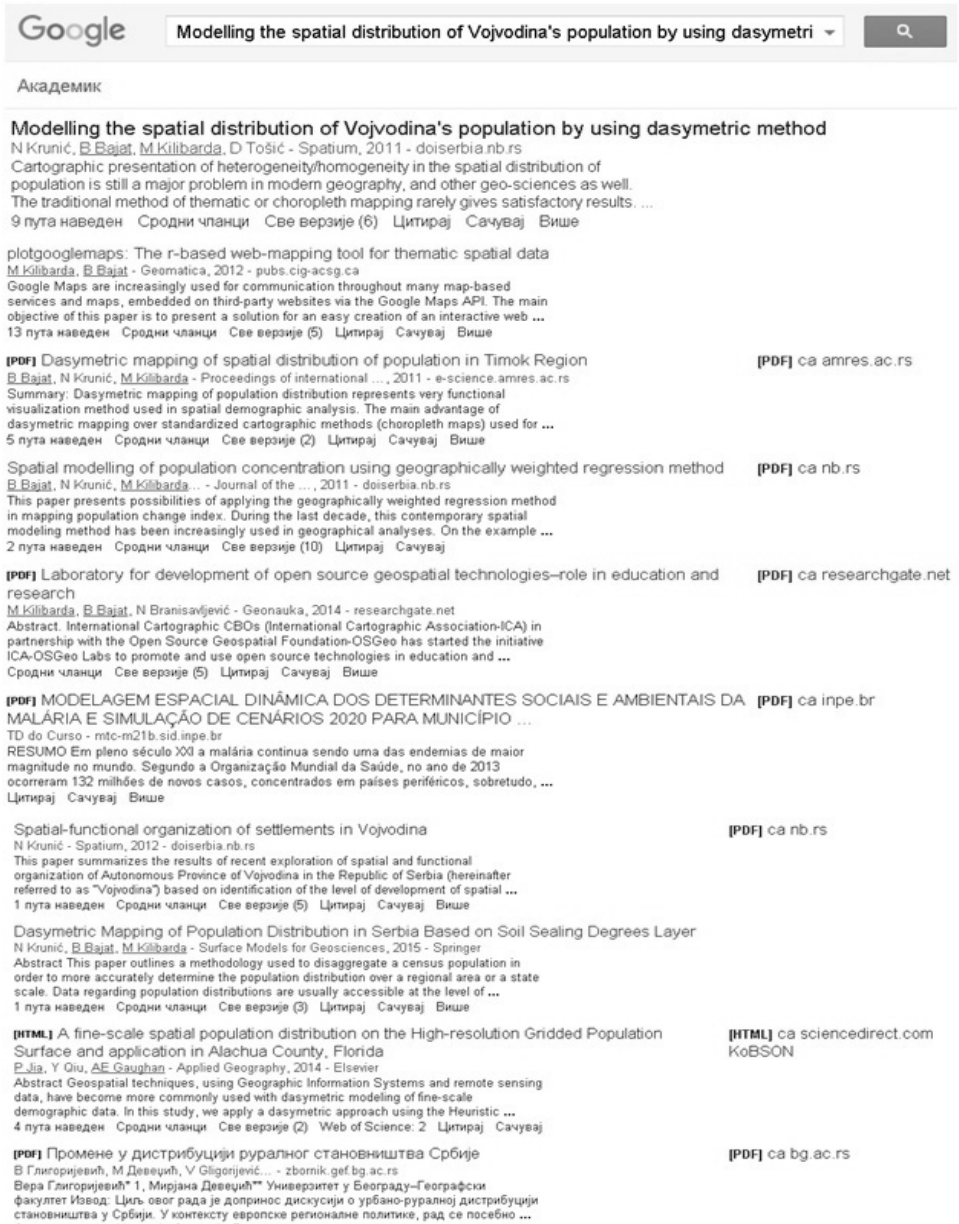


Figure 6. Google Scholar Citation Overview

Figure 7. Scopus's Citation Overview

Considering these examples, it can be concluded that, for the evaluation of the Journal and its ranking, it would be necessary to consult the results obtained from all available sources so that neither the journal itself nor the authors of papers published in it would be harmed by inadequate classification.

6 Categorization of National Journals into the Category of Transport, Urban Planning and Civil Engineering

After having considered the objections to the categorization of national journals for 2013 and pursuant to Article 27 paragraph 1 point 4) and Article 25 paragraph 1 point 5) of the Law on Scientific Research Activities,[21] the Ministry of Education, Science and Technological Development adopted the final list of categorized scientific journals for 2013. The *SPATIUM* journal is the first on the list for the category of Transport, Urban Planning and Civil Engineering for 2013, in the M24 category. The 2013 List is the last official list of categorized national scientific journals that was adopted by the Ministry. The latest proposal of categorization for 2014 is placed on the website of the Ministry of Education, Science and Technological Development.

[21] Закон о научноистраживачкој делатности („Службени гласник РС", бр. 110/05, 50/06-испр. и 18/10))

On the latest, preliminary list of journals, the *SPATIUM* journal is also in the M24 category.

7 Conclusions

The *SPATIUM* journal has been published for 19 years. This analysis covers only a short period from 2009 to 2015 that approximately corresponds to the change of the Editorial Board members and adoption of the *Act on Scientific Journal Editing* 2009 by the Ministry of Education, Science and Technological Development of the Republic of Serbia. The aim of this investigation is to show the role and importance of the *SPATIUM* journal, its multidisciplinary approach to the investigation and publishing of papers in different areas of science.

The analysis of the published texts shows that the number of scientific papers was four times greater than the number of professional papers. Out of the total number of papers, the published scientific papers account for 80 percent, while the professional papers account for 18.62 percent. This supports the fact that *SPATIUM* is primarily oriented towards the scientific communication, although none of its roles in exchanging the experiences among professionals should be neglected.

By constantly struggling for the journal's quality and impact, the Editorial Board of *SPATIUM* contributes to a higher status of the journal. The goal of the Editorial Board is to register the journal for evaluation in the forthcoming period in order to be listed in the Thomson Reuters Master Journal List. By possible entering into the List, the journal would acquire a higher affirmation in the international scientific and professional circles in the fields of spatial and urban planning, architectural design and other related fields, thus also increasing the impact factor of *SPATIUM* as a quantitative measure of its success.

References

Антонић, Сања, Оја Кринуловић and Дејана Каваја Станишић. "Анализа цитираности: Тренутно стање и перспективе". *Панчевачко читалиште*, год. VIII, бр. 15 (2009): 9–11, accessed 19 January 2016, `http://citaliste.rs/casopis/br15/antonic_sanja_l.htm`.

Đurđević, Dušan Ž. "Citation analysis of the University of Priština located in Kosovska Mitrovica". *Sociologija*, vol. 52, number 4 (2010): 421–446, accessed 19 January 2016, `http://www.doiserbia.nb.rs/img/doi/0038-0318/2010/0038-03181004421D.pdf,DOI10.2298/SOC1004421D`, accessed 2016-01-19.

Filipi Matutinović, Stela. *Naučne informacije u srbiji : protok, dostupnost, vrednovanje*. Accessed 19 January 2016, `http://kobson.nb.rs/upload/documents/oNamaPredavanja/PR2013TekstZaDoktorante.pdf`, Belgrade, 2013.

Hussain, Akhtar, Nishat Fatima and Devendra Kumar. "Bibliometric analysis of the 'electronic library' journal (2000-2010)". *Webology*, vol. 8, number 1 (2011): 87, accessed 19 January 2016, http://www.webology.org/2011/v8n1/a87.html.

Jena, Kamal Lochan, Dillip K. Swain, and Sada Bihari Sahu. "Scholarly communication of the electronic library from 2003-2009: a bibliometric study". *The Electronic Library*, vol. 30, number 1 (2012): 103–119, accessed 17 February 2016, http://www.emeraldinsight.com/doi/pdfplus/10.1108/02640471211204097.

Jevremov, Tanja. "Informatička vrednost autorskih ključnih reči u člancima predstavljenim u Srpskom citatnom indeksu". In *Scientific and Professional Conference SNTPI '09*, pages 91–94, 2009. http://ceon.rs/pdf/jevremov.pdf.

Jokić, Maja and Krešimir Zauder. "Bibliometric analysis of rural sociology/sociology and space from 1963-2012". *Sociologija i prostor*, vol. 196, number 2 (2013): 331–349, accessed 17 January 2016, http://hrcak.srce.hr/index.php?show=clanak&id_clanak_jezik=161510,DOI:10.5673/sip.51.2.11.

Ковачевић, Милош. "Категоризација научних часописа: Активност Матичног научног одбора за језик и књижевност (за хуманистичке науке)". *Панчевачко читалиште*, год. VIII, бр. 15 (2009): 4–8, accessed 19 January 2016, http://citaliste.rs/casopis/br15/kovacevic_milos_l.htm.

Martek, Alisa and Snježana Šute. "Bibliometric analysis of the journal arhivski vjesnik in the period 2001-2009". *Arhivski vjesnik*, number 53 (2010): 165–176, accessed 19 January 2016, http://hrcak.srce.hr/62500.

Raičević, Saša. "Database as a source of information on the evaluation of scientific papers in medicine". *Medical Journal of Montenegro*, vol. 1, number 15 (2013): 9–11, 2013. accessed 21 January 2016, http://mjmn.me/11/DatabaseAsASource/DatabaseAsASource.html,doi:10.5937/cma1.2822.

Tella, Adeyinka and Ayotola Aisha Olabooye. "Bibliometric analysis of african journal of library, archives and information science from 2000-2012". *Library Review*, vol. 63, number 4/5 (2014): 305–323, accessed 19 January 2016, http://www.emeraldinsight.com/doi/pdfplus/10.1108/LR-07-2013-0094,DOIhttp://dx.doi.org/10.1108/LR-07-2013-0094.

Thanuskodi, S. "Bibliometric analysis of the journal library philosophy and practice from 2005-2009". *Library Philosophy and Practice*, pages 1–6, 2010. Accessed 19 January 2016, http://unllib.unl.edu/LPP/thanuskodi-lpp.htm.

Tóth, Tibor. "Sastanak s Jacekom Ciesielskim, direktorom izdavačke kuće Versita iz Varšave". *HID-DROBILICA*, vol. IV, number 2 (2007): 2–3, accessed 15 February 2016, http://hidd.hr/wp-content/uploads/sites/178/2013/10/Drobilica_br2_2007.pdf.

Вучковић, Жељко. "Функције часописа у развоју научне комуникације". *Панчевачко читалиште*, год. VIII, бр. 15 (2009): 2–3, accessed 14 February 2016, http://www.citaliste.com/casopis/br15/Citaliste_br15.pdf.

Vukasović, Predrag. "Analysis of citations, origins, purposes and basic principles". *Strani pravni život*, number 1 (2009): 211–230.

Web pages

"DOI Serbia – Naučni časopisi objavljeni u Serbia u punom tekstu" /DOI Serbia – Scientific journals published in Serbia in the full text/, The European Library, accessed 15.02.2016, http://www.theeuropeanlibrary.org/tel4/collection/a0155?id=a0155&locale=sr

"Kategorizacija časopisa" /Categorization of Journals/, Ministry of Education, Science and Technological Development, accessed 14.02.2016, http://www.mpn.gov.rs/casopisi/

SCIndex – Serbian Citation Index, accessed 15.02.2016, http://scindeks.ceon.rs

"Spatium" DOISerbia, accessed 15.02.2016, http://www.doiserbia.nb.rs/journal.aspx?issn=1450-569X

"Spatium" SCIndex – Serbian Citation Index, accessed 15.02.2016 , http://scindeks.ceon.rs/journaldetails.aspx?issn=1450-569X

"Spatium, 1450-569X " Academic Journals Database – Disseminating quality controlled scientific knowledge, accessed 15.02.2016, http://journaldatabase.info/journal/issn1450-569X

"Spatium, 1450-569X (Print)" Directory of open access journals, accessed 15.02.2016, http://doaj.org/toc/d3f460f04f2f4f849aa9feef69cec258

"SPATIUM-The Journal of Institute of Architecture and Urban & Spatial Planning of Serbia" de Gruyter Open, accessed 15.02.2016, http://www.degruyter.com/view/j/spat

http://scindeks.ceon.rs/Related.aspx?artcit=1450-569X1124045K

http://scholar.google.com/scholar?cites=12858484065582622O5&as_sdt=2005&sciodt=0,5&hl=sr

E-books and new dimension of reading

UDC 004.738.5:655.41
DOI 10.18485/infotheca.2016.16.1_2.5

Gordana Nedeljkov
gordana.nedeljkov06@gmail.com
The Andrić Institute, Višegrad

ABSTRACT: Technology development contributed to everyday growth of e-books. Access to e-books is provided by editors and distributors of electronic editions, for free or in return for payment. In Serbia there are a couple of editors that are publishing e-books and the increase of digital libraries that provide access to e-books is also notable. E-books can be read on computers, laptops, tablets and smart phones, but e-book readers are the most suitable for longer reading. Easy access, availability 24 hours a day, efficiency in terms of cost and space and other advantages of e-books have contributed to the increase of the reading public.

KEYWORDS: digitization, e-book, e-publishing, e-book readers, reading

PAPER SUBMITTED: 25 March 2016
PAPER ACCEPTED: 07 May 2016

1 Introduction

During its many centuries of life, book has changed its physical appearing, but not its purpose. It has always served as an instrument of documenting, keeping and transferring knowledge and information. According to UNESCO's definition book is "non-periodical printed publication having 49 or more pages which do not include cover".[1] In Encyclopedia Britannica it is specified that a book is an instrument of communication, whereby the book is observed as a written or printed message of certain length, meant for public usage and recorded on material that is durable enough to provide its easy transfer.[2]

Today, books are mostly present in printed form, but their appearance in digital form is also very common. E-books are mostly used by new generation, but elder

[1] Business Dictionary: `http://www.businessdictionary.com/definition/book.html`, accessed 14 April 2016.

[2] Encyclopedia Britannica, accessed 21 March 2016, `http://www.britannica.com/EBchecked/topic/73295/book.`

users, especially those who are familiar with usage of computers and new IT, show interest in reading e-books too.

Some of the questions that will be discussed in this paper are: how can we have access to e-books, what are the differences in reading electronic and printed forms of books and what are advantages and disadvantages of e-books compared to the printed editions. Besides that, themes such as e-publishing and digital libraries in Serbia and over the world will be discussed, as well as devices for reading e-books and changes made in the reading process itself.

2 Term and emergence of e-books

Electronic books include monograph publications that exist at the same time on paper and in digital form or only in digital form. Development of technology contributed to everyday increase of electronic books. E-books are mostly distributed on compact disks or over the web, and they could be read on computers, mobile devices or so-called e-book readers. The first idea that followed development of e-books was that readers could leave comments on text and communicate with authors (Васиљевић, 2006). Though, as the author of above-mentioned paper quotes, the idea was not adopted, e-books continued to exist having the same function as printed books – as medium for transfer of knowledge and information, but now in a faster and easier way. The development of e-books changed reading process itself and positive aspects of those changes were recognized in literature and confirmed in practice, which caused increase of e-books popularity.

In the English language there are two terms related to e-books that explain their emergence: "born-digital" and "made-digital". Term "born-digital" is used for books that are originally made in digital form. In March 2000 the first book in e-form only was published: novel "Riding the Bullet" by American author Stephen King (Трифуновић, 2005). The second term, "made-digital", refers to books which are converted from printed to electronic form. Digitization or "converting a book from printed into machine-readable format, using special hardware and software solutions" (Вранеш and Марковић, 2008) involves a complex set of tasks that include material selection, digitization of printed publication, processing and marking text so it could be searchable and preserving book in a medium or on server in a specific format. Common formats for electronic documents like ASCII, HTML and PDF can also be used for e-books, but there are specialized e-book formats such as ePub, Mobi, Nook, etc.

3 How to get e-books?

3.1 Digital libraries and publishing of electronic books

Books in digital form can be found as a part of digitization project, in digital libraries around the world. Digital library is a collection of digital objects of certified quality that are, through using of new technologies, made or collected and organized according to international principles and made available for users (IFLA/UNESCO, 2010). Considerable number of publications in digital libraries are in open access which implies that "every user who has right of access to the Internet has right to read, download, print and use digital content that is in open access, with only obligation to quote it properly" (Филипи-Матутиновић, 2013). It is practice around the world that author agrees to open access to his work by signing license. OA (Open Access) databases allow free access to full text e-books, and some of them are: Project Gutenberg, Google Books, Europeana, Open Library, Internet Archive, Digital Public Library of America, Daily Lit, Classic Reader and National Academic Press.

Publishers and distributors of electronic editions also provide access to e-books, for free or in return for payment. E-books can be purchased through subscription or through license, whether it is one book or a collection. The most known commercial providers are: Questia, netLibrary, ebrary, Oxford Reference Online, Springer, Elsevier, ScienceDirect, EBSCO and Emerald.

The level of availability and terms of access depend on providers and copyright. Publishing electronic books copyright has to be respected, as well as in the case of classic printed edition. Companies that are specialized in publishing e-books often publish editions that are no longer under copyright. If that is not the case, publishing e-books is defined by copyright law of country in which publication is being published. Most countries protect foreign authors and copyright holders, if they are from a country which is signatory of an international agreement in copyright issue. One of such agreements is Berne Convention, global arrangement about copyright protection on the Internet, according to which national laws of each country provide an automatic protection of copyright works, pursuant to the law of the contracting party. Serbia is one of the contracting parties of the Berne Convention, and further information on copyright in Serbia can be found in Copyright and medium-sized law.

3.2 E-books worldwide

The first project that provided access to electronic books was Project Gutenberg, with mission to encourage creation and development of electronic books.[3] Michael

[3] Project Gutenberg: `http://www.gutenberg.org/`, accessed 23 March 2016.

Hart was initiator of this project and in 1971 he started to digitize texts that were no longer under copyright and to upload it on servers and later on the Internet. Over time, number of digitized books increased and now there are more than 50,000 digitized books on project's website. All books that are included into Project Gutenberg are in full text and completely searchable. On the home page of Project Gutenberg's website is stated that usage of digitized books is in accordance with the Copyright law of the United States and users from other countries are obliged to check copyright law in their own country before the download and distribution of books.

Project of digitization of European cultural heritage was named Europeana[4] and was launched with an idea to provide free access to digitized content, protection of cultural heritage and exchange of ideas and information. Portal Europeana provides access to 24 million digital documents, among which are electronic books. Numbers of partners and amount of accessible contents increase every day.

Google books[5] provide searching and access to books in electronic form. Visibility of obtained search result and its availability depends on the copyright. If copyright has expired or given to company Google books and the book is uploaded as a public good, it is possible to download the whole book. In other cases, authors or copyright holders define the level of book availability. Within this project there is service Google Books Library Project which as a search result gives bibliographic data and if the book is under copyright links to internet bookstores where book can be bought and libraries where it can be lent.

Open Library[6] is a database that contains links to more than million e-books. Certain number of books is in open access, and there is also an option for lending books from the online library.

Databases that were mentioned contain various electronic books, so besides works of classic and modern literature books from different scientific fields can be found. In that way, by putting digital content online, users from all over the world have quick and easy access to literature that they need for study, research or reading for fun.

3.3 E-books in Serbia

Today a great number of Serbian institutions of culture and education have their own digital library or digital repository online and provide access to books in electronic form (IFLA/UNESCO, 2010). These databases are mostly formed by digitization of library's collections or by putting electronic editions of this institution online and can be accessed free of charge.

[4] Europeana, `http://www.europeana.eu/portal/`, accessed 14 April 2016.
[5] Google books, `https://books.google.com/`, accessed 11 April 2016.
[6] Open Library, `https://openlibrary.org/`, accessed 11 April 2016.

The first undertaking launched with aim to provide access to books in digital form in Serbia was Project Rastko, as a Serbian version of Project Gutenberg. Project Rastko is a sort of an online library formed of books from scientific fields and art that belong to Serbian or similar cultures, with the aim to make cultural heritage accessible to national and international public. On website of the project is stated that this project is "non-profit, non-governmental and voluntary and is realized through voluntary activity of individuals and institutions that are engaged in research, development and preservation of Serbian culture"[7] and also that digitization of books was conducted pursuant to copyright regulations. Idea of project establishment originated in 1994 and in September 1997 project started with public work. Digital library Rastko today contains a large number of books in electronic form, in open access and with fully searchable text. It is stated on the project's website that extent of usage of author's content is determined by permits of copyright holders, defined by written agreement compatible with national and international copyright law regulations. Over time, Project Rastko has been expanded and today it includes other countries that digitize and put online electronic publications that refer to Serbia or Serbian culture or are directly related to it. In that way the whole project is enriched and popularized outside of Serbia. As a part of the project there are: Project Rastko Timisoara: the library of Serbian culture in Romania, Project Rastko Budapest-Szentendre: the library of culture and tradition of Serbs in Hungary, Project Rastko Kiev-Lviv: the library of Ukrainian-Serbian cultural relations, Project Rastko Bulgaria: the library of Bulgarian-Serbian cultural relationship, and others.

The Digital library of Faculty of Philology, University in Belgrade,[8] has been created by digitizing Faculty's editions, and currently consists of more than 107,000 pages of digitized material of monograph and periodical publications. Library was formed as support of distance learning and is intended for students, teachers and researches and its content is searchable by author, title, professional and subject classification and by every word in text.

The Digital National library of Serbia[9] consists of a few books collections, periodicals, photo documents and other materials from the fund of the National library of Serbia, with the aim to provide open access to knowledge and information. Ser-

[7] Project Rastko: the library of Serbian culture, `http://www.rastko.rs`, accessed 22 March 2016.

[8] Digital library of the Faculty of Philology, `http://www.fil.bg.ac.rs/lang/sr/biblioteke/digitalna-biblioteka/`, accessed 21 March 2016.

[9] Digital National Library of Serbia, http://www.digitalna.nb.rs/, accessed 21 March 2016.

bian children's digital library is a part of this library, with 127 books for children that were digitized as part of the project International children's digital library.[10]

Project Anthology of Serbian literature[11] has the aim to provide easy access to the most famous works of Serbian literature for all interested readers around the word. All books were digitized according to the copyright regulations. Book selection was conducted by the Faculty of Teacher Training and digitized by Microsoft Center for Software Development in Belgrade. Among 130 free accessible books there are works of Radoje Domanović, Laza Kostić, Desanka Maksimović, Jovan Sterija Popović and other Serbian authors. Thanks to the electronic version of these titles which are used as school reading faster and simultaneous access to the publication is provided for a large number of users and the problem of insufficient number of copies of the book is solved.

Access to electronic contents provided by some of the commercial providers in Serbia is enabled by Serbian Library Consortium for Coordinated Acquisition (KoBSON) that was formed in 2002 as one of the first initiatives for providing access to scientific information in digital form.[12] KoBSON was initially focused on purchase of electronic periodicals and later introduced the possibility of access to books in electronic form. Access to electronic publication via KoBSON has its restrictions: it can be only accessed through academic network, from institutions that were founded by the Government of Republic of Serbia and that are under the jurisdiction of the Ministry of Culture or, in particular circumstances, from home. In 2016 thanks to KoBSON users from Serbia can access almost 160,000 electronic books from different scientific areas provided by Springer, EBSCO eBook, Google Books, Pub Med, ScienceDirect, Cleveland Medicine Index, DOAB and Hein on Line.

Although the e-book publishing market in Serbia is still not developed enough, there is some progress concerning that issue so today we have a couple of publishers that publish books in electronic form. One of the first publishers of e-books only in Serbia was NM Libris, founded in 2013. It functioned as follows: part of the books was accessible for free to everybody, and by creating a user account and paying membership fee user was able to access and download all the books that were put on the website. Besides that, users could create and organize their own library. At present, website of this publishing house is out of function for unknown reasons.

The publishing house Booka joined to the Serbian publishers of e-books with two books: "Below deck" by Vladimir Arsenijević and "39 songs" by Nebojša Kri-

[10] International Children's Digital Library, `http://en.childrenslibrary.org/`, accessed 22 April 2016.

[11] Project Anthology of Serbian Literature, `http://www.antologijasrpskeknjizevnosti.rs/`, accessed 23 March 2016.

[12] Serbian Library Consortium for Coordinated Acquisition, `http://kobson.nb.rs/kobson/`, accessed 21 March 2016.

vokuća, which are uploaded on couple of websites specialized for selling e-books. Booka announced publishing of other books in digital form.[13] On the 15th of May 2014 publisher Arhipelag announced launching of e-publishing in series e-Arhipelag, saying that book is "a text that we read and it is much more important than form in which that text comes to the readers".[14] For now, eight e-books have been published and it was planned to keep publishing other titles. Company Media Art Content DOO from Novi Sad also publishes e-books that can be bought via international commercial distributors.[15]

Besides mentioned ones, there are other publishers in Serbia that expand its publishing from printed to electronic editions; however, e-publishing in Serbia is not fully established. For example, an attempt of search of co-operative catalogue of Virtual library of Serbia using basic and advanced search, looking for information about books published in digital form is not applicable. It is not possible to collect needed data that way, because in basic and advanced search in Virtual library of Serbia catalogue there is no option of searching only books in e-form. The only possibility is the search within expert search during which is needed to make special complex query. By searching co-operative catalogue with a query EA="*"/mon we get 10,648 books in e-form, among which publications in foreign languages, PhD theses and digitized old books. Trying to get information about e-publishing in Serbia for past 15 years, we may narrow the search and use query (EA="*"/mon and LA=(scc or srp) and PY=20*) not TI=doktorsk*. In that way we get 1,690 results, i.e. monographs in electronic form published in Serbian language from 2,000 until now. Results of this search include, besides e-books, catalogues, guidebooks and other documents in e-form, and links given with every result lead to publications that are on web. By this search it is possible to gain insight into state of publishing activities in Serbia and monitor its development.

4 Devices for reading e-books

E-books have some requirements concerning hardware and software in order to be used. In paper "Electronic publications and libraries – an overview" while speaking of e-books author states that there is "always present some kind of resistance to this kind of publications, as the technology is still not at the level that could

[13] Publishing House Booka, `http://www.booka.in/knjige/ebooks`, accessed 21 March 2016.

[14] Publishing House Arhipelag, `http://www.arhipelag.rs/arhipelag-magazin/gojko-bozovic-magazin-3/`, accessed 22 March 2016.

[15] Media Art Content LTD, `http://www.elektronske-knjige.org/`, accessed 21 March 2016.

provide longer and more comfortable reading" (Васиљевић, 2006). With technology improvement, situation has changed. E-books can be read on computers, laptops, tablets or smart phones and those devices can be used for shorter reading. For longer reading it is recommended to use devices that are at the first place conceived to be readers of electronic books. With increasingly advanced electronic readers, reading e-books is not very different from reading printed edition. Those devices provide keeping a large number of books, which can be carried out and shared with others in every moment. E-books and readers are mostly used by researchers and scientific workers, who have become aware of advantages of this kind of reading and research work (Nehring, 2013). For the people who spend most of their time travelling, it is more convenient to put books they need in an e-book reader than to carry around several printed books.

Depending on manufacturers and models, e-book readers have different characteristics. There are simple models that allow transfer of e-books from computers, forming a personal library, searching text and making notes. E-book readers are connected to specialized stores via Internet, in which users can buy books, usually for a lower price that printed version. Besides simple ones, there are advanced e-book readers that have characteristics like tablets or smart phones. These kinds of devices have additional applications among which some are not intended for reading (i.e. application for listening radio or playing games). It has been shown in practice that these advanced devices are not the best choice for reading e-books because, focusing on additional characteristics, they have lost quality when it comes to what they should be in the first place – medium that will provide reading book in e-form as simple as possible. Devices that users have rated as the most suitable for longer reading support the so-called e-ink technology. This technology gives the impression of reading from paper, so during reading eyes are not getting tired as it happens while reading from monitor or tablet. Also, this screen allows reading in the dark and in daylight, with anti-reflective coating and anti-flicker technology. Advantage of the e-ink technology is better battery capacity, which is not the case with devices that do not use this technology. There are devices with regular screen and those with touch screen display.

Today e-book readers have various forms and characteristics, so users can choose the most suitable one among diverse models. Regarding the supply of devices for e-books reading in Serbia, certain models with different specifications and various prices can be found, but the offer is far from diverse. The other possibility is buying an e-book reader in a store abroad. As an alternative to purchase of e-book reader there is a possibility of reading from computer, tablet or a smart phone. There are a large number of applications that provide reading from mentioned devices and also

make manipulation with books and text among them possible.[16] These applications make reading from a computer easier, but for the most comfortable reading of books in electronic form e-book reader is more suitable.

5 Advantages and disadvantages of e-books and the new approach to reading

Scientists and researchers seek for a quick and easy access to information and e-books provide exactly that. New generations expect to find all needed literature online (Adema, 2012). Accessibility 24 hours a day is one of the advantages of electronic editions opposite to the printed ones. Publications in electronic form that are uploaded online become immediately available to everybody. Unlimited number of users can read one e-book at the same time, not having to be physically present in a library. In that way, problem of not having enough number of copies of a book is solved.

Users choose to read e-books because these editions do not need a lot of storage place. With appearance of mobile devices for reading, or so-called e-book readers, reading gains new dimension. Now it is possible to take a large amount of books with you. Besides, the text in e-form is searchable,[17] letter size and font can be edited according to personal taste and hyperlinks can lead from one page to another (Трифуновић, 2005). Text marks can be made in e-books and some of e-book readers have built-in electronic dictionaries.

E-books can be quickly exchanged, copied and transferred from one device to another. One of the advantages of e-books is its economy and environmental friendliness, as they do not use paper at all. Digitization of printed editions provides preservation and protection of old and rare books and other publications from frequent and intensive usage. Besides, electronic editions of old and rare books sometimes are the only way to present national and international cultural heritage to public.

Disadvantages attributed to e-books refer more to devices for e-books reading, than to e-books themselves. One of those disadvantages is the inability of comfortable reading during longer period of time. Reading books from computer is difficult because of disturbing factors such as other running applications, windows and similar things that draw attention from reading process itself (Pettigrew, 2015). By moving from computer monitor to e-book readers, where disturbing factors are removed or at least reduced, these disadvantages are slowly disappearing.

[16] Some of these applications are: Wattpad, Amazon Kindle, FBReader and Kobo eBooks.

[17] The most used format for e-books encoding that enables the full-text search is the international format TEI (Text Encoding Initiative).

Depending on manufacturers and models e-book readers support only some of the formats and that can be considered as a limitation for using e-books. However, various applications that read different e-book formats can be found on the Internet, so this disadvantage is practically insignificant. Applications such as Calibre[18] and Sumatra PDF[19] that are very simple for use and possibility of converting e-book in different formats help users to create their own digital library and manipulate with books in it. Batteries for e-book readers have a large capacity so they can be used for many hours.

Appearance and usage of e-books influenced reading process itself. Many authors claim that e-books can be effectively used for promotion of reading (Weber and Cavanaugh, 2006) and information literacy for children of school age (Schiemann, 2016). Thanks to the Internet the books are nowadays available more than ever and it is considered that this can contribute to the "production" of readers who will read different types of publications (Pettigrew, 2015) and the increase of overall reading public.

Although e-books were originally designed as a direct link between authors and readers, they contributed to the development of other types of interaction. Innovations in technology helped the development of so-called interactive electronic books that provide high level of interaction among user, text in electronic form and digital environment (Bozkurt and Bozkaya, 2015). In addition to creation notes in text, e-book readers provide the opportunity of fast content sharing through the Internet and social networks, exchange of thoughts and ideas and reflections on it. That is how immediate exchange of information on the read text is accomplished, and thanks to the following the content created by others the readers are allowed to find appropriate literature for reading and research in the simplest way possible (Софронијевић, 2015).

Books in electronic form contributed to the connection between reader and text that is being read. Authors define electronic books as "adjustable" and as its characteristics they state possibility of text size adjustment, underlining text, writing notes, interactive dictionaries and software for reading out loud.[20] Large memory of e-book readers provides users the possibility to take their own library always with them, instead of being limited to only couple of books that they can use at every moment. Thanks to the advanced technology that devices for reading e-books use,

[18] Calibre e-book menagement: `http://calibre-ebook.com/`, accessed 22 March 2016.

[19] Sumatra PDF, `http://www.sumatrapdfreader.org/free-pdf-reader.html`, accessed 22 March 2016.

[20] Software for reading e-books out loud are adjusted to English language. Some software also allows reading out loud in Serbian, but it is still not at satisfactory level: words are not clearly pronounced and it can hardly be followed.

they have become practical for everyday reading, because they can be used in various conditions.

6 Conclusion

Given the fact that we live in a time where technology is present in almost every part of life, its influence in activities such as reading was expected. Digital environment contributed to the change in reading process by enabling connections in networks of knowledge and information, reciprocal communication among readers and closer relationship with text that is being read. Growth of e-books, its acceptable price, easier and simpler access and more advanced electronic devices for reading, will contribute to its ever-bigger distribution and usage. Advantages such as simpler access, 24 hours a day availability and efficiency in terms of cost and space, will influence the growth of number of people who read e-books, which will also increase growth of reading public altogether.

References

Adema, Janneke. "On open books and fluid humanities". *Scholarly and Research Communication*, vol. 3, number 3 (2012):16, accessed 12 April 2016, `http://src-online.ca/src/index.php/src/article/view/92/107`.

Bozkurt, Aras and Mujgan Bozkaya. "Evaluation criteria for interactive e-books for open and distance learning". *The International Review of Research in Open and Distributed Learning*, vol. 16, number 5 (2015):58–82, accessed 24 April 2016, `https://www.academia.edu/16304364/Evaluation_Criteria_for_Interactive_E-Books_for_Open_and_Distance_Learning`.

Филипи-Матутиновић, Стела. *Ауторска права за библиотекаре: приручник*. Београд: Универзитетска библиотека „Светозар Марковић", 2013.

IFLA/UNESCO Manifesto for Digital Libraries, 2010. accessed 13 April 2016, `http://www.ifla.org/publications/iflaunesco-manifesto-for-digital-libraries`,

Nehring, Daniel. "How Useful are E-readers for Academic Reading, Really?" accessed 24 March 2016, `http://www.socialsciencespace.com/2013/03/how-useful-are-e-readers-for-academic-reading/`, 2013.

Pettigrew, Sarah Ellen. "The future of the book and the slow book movement" in *History of the book: Disrupting Society from Tablet to Tablet*, 139–159. Western Oregon University, 2015, accessed 23 March 2016, `http://digitalcommons.wou.edu/cgi/viewcontent.cgi?article=1014&context=history_of_book`.

Софронијевић, Адам. „Нова парадигма сарадње у библиотекама", Докт. дисертација, Универзитет у Београду, Филолошки факултет, 2015.

Schiemann, Christopher. "E-books Can be a Great Choice for Middle Schoolers", in School Library Journal, February 2016, accessed 14 April 2016, http://www.slj.com/2016/02/students/ebooks-can-be-a-great-choice-for-middle-schoolers/.

Трифуновић, Богдан. „Електронске књиге". *Глас библиотеке*, бр. 12 (2005): 15, accessed 21 March 2016, http://www.cacak-dis.rs/elektronska_izdanja/Glas_biblioteke_12/pdf_gb_12/Bogdan%20Trifunovic-Elektronske%20knjige.pdf.

Васиљевић, Наташа. „Електронске публикације и библиотеке – пресек стања". *Инфотека*, год. 7, бр. 1-2 (2006):75–80, accessed 22 March 2016, http://infoteka.bg.ac.rs/pdf/Srp/2006/INFOTEKA_VII_1-2_Jun2006_75-80.pdf.

Weber, Christine L and Terence W Cavanaugh. "Promoting reading: Using ebooks with gifted and advanced readers". *Gifted Child Today*, year 4, vol. 29 (2006):56–63, accessed 23 March 2016, http://files.eric.ed.gov/fulltext/EJ746309.pdf.

Вранеш, Александра и Љиљана Марковић. *Од рукописа до библиотеке : појмовник*. Београд : Филолошки факултет, 2008.

Preparation of Multimedia Document "YU Rock Scene"

UDC 004.55:378.147]:02(497.11)
DOI 10.18485/infotheca.2016.16.1_2.6

ABSTRACT: This study presents the preparation process of the multimedia document entitled "YU ROCK SCENE" in which participants were senior students of undergraduate studies of the Department of Library and Information Science at the University of Belgrade, Faculty of Philology during the academic year 2014/2015, as a part of the subject Multimedia Documents. This study gives an overview of the historical development of rock and roll in the territory of the former Yugoslavia, rock scene in Yugoslav republics, as well as the influence of rock music on other arts and media. Special attention was given to the technical implementation of the project, including stages such as planning, processing, designing and creation of the multimedia document itself.

KEYWORDS: multimedia document, library science, information science, rock and roll, music, Yugoslavia

PAPER SUBMITTED: 28 March 2016
PAPER ACCEPTED: 06 May 2016

Milena Obradović
milenaacm1@hotmail.com

Aleksandra Arsenijević
aki.7424@gmail.com

Mihailo Škorić
miccersoft@gmail.com

University of Belgrade, Faculty of Philology

1 Historical Development of Rock and Roll in the Territory of Former Yugoslavia

Rock and roll in the territory of former Yugoslavia is rooted in the 1920s when new instruments, such as saxophone and guitar, began to emerge (Fajfić and Nenad, 2009, pp. 18–21). The emergence of these instruments influenced Yugoslav musicians to start playing jazz, in accordance with the world trends. Musicians who played the said new instruments and jazz rhythms before World War II were rare, while they were completely obsolete during World War II. Pupil and student orchestra that regularly played jazz at dance parties began to be formed only after the end of World War II. Some people believed at the time, under the influence of media and

war, that jazz was "devil's music" causing the youth to become rebellious and that it should be banned. In spite of these attitudes, jazz persisted, which was also a result of regular listening to (jazz) music, including the world hits of the time, broadcasted by *Radio Luxembourg*. Still, during the 1950s, the so-called popular music had much greater presence.

Due to the conflict with the Soviet Union during the Cold War, Yugoslavia, one of the founding countries of the Non-Aligned Movement, was more open to the West and all products of pop-culture, especially American pop-culture. Yugoslavia was thus the only Communist country that participated in the *Eurovision Song Contest* in 1961. At the beginning of the following decade, musicians, and especially youth, began following primarily British and American trends. Gramophone record stores were becoming more and more prevalent, and in particular places music could be listened on jukeboxes. Production of Coca-Cola began in this period in these territories, dressing style was being adapted to the trends in the West, and the emergence of jeans caused a genuine revolution in the fashion world. Rock and roll too was not lagging behind these trends. The musicians of the time started making cover versions of world famous music hits, especially those of Elvis Presley, Chuck Berry, Buddy Holly and many other authors.

It is difficult to decide who was the most deserving for the breakthrough of rock and roll into the Yugoslav music scene. Many people believe that Mile Lojpur and Perica Stojančić from Belgrade and Karlo Metikoš from Zagreb played the most significant role in this undertaking (Fajfić and Nenad, 2009, стр. 18–21). As a music genre, rock and roll spread very quickly across the entire Yugoslavia. It was then that two musical directions developed in rock and roll music: country pop rock (combination of ethno sound and rock style) and the western, mostly Anglo-American rock. Unlike the world rock scene whose texts were a powerful means of propaganda against wars and class conflicts, original domestic rock texts had visual and musical identity related to youth fantasies, dreams about success, as well as to the issues of the generation.

Number of rock groups and solo performers was increasing, and the 1970s are considered the most fruitful years, since they brought mass production and great popularity to the rock scene. Bands that were established in the beginning of the 1970s were greatly influenced by the hippy movement, but then branching of rock music into sub-branches began, giving rise to hard rock, progressive rock, art rock, jazz, blues and others. At the end of that decade, punk appeared (under the influence of British and American punk bands such as *The Sex Pistols*, *The Clash*, *The Stooges*, *Ramones* and others), as well as the New Wave as a completely original cultural and artistic movement.

Punk itself represents an antithesis of the established political, social and economic principles of the society, while the New Wave was characterized by a special

rebellion, change in mindset and the disposition of the Yugoslav society which became more open and more critical, with more open understandings and tastes in the single-party political system of Yugoslavia. Until the emergence of punk and the New Wave the main topic of lyrics was love. With the development of punk and the New Wave the texts gained new breadth and complexity which was at variety with the social, cultural and political norms of the time. Although the New Wave was equated with trashy music (Kostić, 2013a), today it is very reputable and it is considered that it was a phenomenon that was never repeated.

Almost the entire following decade was marked by the New Wave, and near the end of the 1980s new groups emerged which were greatly influenced by their older colleagues. Some performers only remained on the scene for a short time, while on the other hand many of them are more or less still active today. Each of them left an indelible mark on the Yugoslav rock scene and great influence on the society and culture in general. In Yugoslavia, as well as in the world, rock music represented a comprehensive movement that influenced numerous events.

2 YU Rock Scene per Territories

In this part of the Multimedia Project, many bands and performers are represented from the territory of the Former Yugoslavia, i.e. from Serbia, Croatia, Slovenia, Bosnia and Herzegovina, Macedonia and Montenegro. In short biographies of groups and performers, the most important years, members and events in which they participated are listed, as well as their complete discography.

2.1 YU Rock Scene in Serbia

Belgrade was the Serbian centre for establishment, operation of a large number of bands and affirmation of solo performers. During the 1960s, the following bands appeared: *Zlatni dečaci*, *Elipse*, *Korni grupa*, *Porodična manufaktura crnog hleba*, *Dogovor iz 1804.*, *Iskre*, *Crni Biseri*, Duet *Vlada i Bajka*, *Džentlmeni*, *Plamenih 5* and *Siluete*. Solo performers also appeared on the scene, some of whom were members of some of the said music groups, such as Mile Lojpur, Perica Stojančić, Dušan Prelević, Kornelije Kovač, Zoran Simjanović, Slobodan Boba Stefanović, Đorđe Marjanović and many others. Apart from concerts and festivals that emerged at that time, such as *Beogradsko proleće* 'Belgrade Spring' or *Beogradska gitarijada* 'Belgrade Guitar Fest', the most important places for going out were clubs in which rock and roll was played. Clubs such as *Gradski podrum*, *Dom omladine* and *Euridika* were favourite places where youth gathered. During the 1970s, bands and performers emerged whose music was particularly influenced by the hippy movement. The following groups stood out: *YU grupa*, *Suncokret*, *Bulevar*, *Poslednja igra*

leptira, *Zana*, *S vremena na vreme*, *Generacija pet*, *Riblja čorba*, *Igra staklenih perli*, *Laki Pingvini*, and the following solo performers emerged: Lazar Ristovski, Srđan Marjanović, Oliver Mandić, Dejan Cukić, Slađana Milošević, Aleksandar Sanja Ilić and Rambo Amadeus. Groups primarily influenced by the New Wave appeared on the big Belgrade rock scene in the early 1980s. These groups included *Idoli*, *Šarlo akrobata*, *Električni orgazam*, *Ekatarina Velika*, *Disciplina kičme*, *Partibrejkers* and *U Škripcu*. The following rock groups then appeared: *Piloti*, *Viktorija*, *Džakarta*, *Ruž*, *Divlji Anđjeli*, *Alisa*, *Oktobar 1864.*, *Sirova koža*, *Bajaga i instruktori*, *S.T.R.A.H.*, *Van Gog* and *Vampiri*. Solo performers who managed to stand out include: Vladimir Vlada Divljan, Bebi Dol, Milan Delčić Delča and Toni Montano.

Although Belgrade was epicentre of Serbia concerning cultural events and creation of rock music, it should be pointed out that the rest of Serbia was not lagging behind. Larger cities in Serbia significant for the Yugoslav rock culture, which gave birth to large rock groups include: Niš, Čačak, Kragujevac, Novi Sad, and several smaller towns, such as Bečej (town in which Eva Braun originated) and which made history of the Former Yugoslavia with one or two bands. Important groups originated in Niš – *Galija*, *Kerber* and *Dobri Isak*, and during the existence of Yugoslavia, large rock concerts and performances were held there. During that time, 84 rock and roll groups were established in Čačak, and most of them originated from the Cultural Artistic Society *Abrašević*. The first and the most popular group in Čačak in the 1960s was *Bele višnje*, and the following new groups appeared after them: *Zvečarke*, *Crne mambe*, *Srebrne senke*, *Dečaci sa Morave*, *Hermelini*, *Čačanski plemići*, *Safiri*, *Nautilus*, *Kućni savet*, *Sebastijan*, *Moris*, *Mračni Mo*, *Spin*, *Visoki napon*, *Krvna grupa*, etc. Two top-quality rock performers originated in Čačak: Radomir Mihailović Točak and Borisav Đorđević – Bora Čorba. With time, Kragujevac became synonym for the group *Smak*, one of the oldest and most significant rock bands in these territories. Other bands which originated in Kragujevac also include band *Osvajači* and Punk group *KBO!*. One of the most important representatives of Novi Sad is Đorđe Balašević, who was member of bands *Žetva* and *Rani Mraz*, before his successful solo career. The group *Garavi Sokak* is also from Novi Sad.

2.2 YU Rock Scene in Croatia

First wave of rock and roll in Croatia emerged in the closing years of the 1950s. However, first genuine rock bands, especially from Zagreb, Pula, Rijeka and Split were established in the beginning of the 1960s, and the most famous of them include: *Bijele strijele*, *Grupa 220*, *Crveni koralji*, *Delfini*, *Bezimeni*, *Atomi*, *Kvartet 4M* and many others. One of the first great and world famous Croatian solo performers, who is also considered especially deserving of representing and spreading rock music in these territories is Karlo Metikoš, better known as Matt Collins. Apart from him, Ivica

Percl, Josipa Lisac, Zdenka Kovačiček, Drago Mlinarec and other authors began to develop their successful careers. The most important bands in Croatia in the 1970s, some of which belonged to the New Wave, include: *Azra*, *Haustor*, *Parni Valjak* and *Prljavo Kazalište*, *Atomsko sklonište*, *Aerodrom*, *Tajm*, *Paraf*, *Drugi način*, *Metak* and many others. The most important punk rock bands of that period in Croatia include *Paraf* and *KUD Idijoti*. Some of the members of these bands later continued their very successful solo careers, e.g. Dado Topić, Dino Dvornik, Branimir Džoni Štulić, Darko Rundek and Masimo Savić. Among others, *Psihomodo pop*, *Neki to vole vruće*, *Dorijan Grej*, *Osmi putnik*, *Đavoli* and *Fit* appeared on the Croatian rock scene in the 1980s. In the ending years of the 1980s and in the beginning of the 1990s, Croatian music scene became real industry that produced very popular bands.

2.3 YU Rock Scene in Slovenia

Chansons and folk music were always the most important music forms, however in the 1960s, as well as in all other parts of the Former Yugoslavia, under the western influence, an increasing number of bands started to emerge in Ljubljana, Kopar, Maribor and Kranj, and that is when the real cultural rock revolution commenced. Special radio shows were made and articles were written which were dedicated to this new music genre. One of the first Slovenian rock bands which made a breakthrough to the big rock scene of the entire Yugoslavia was group *Kameleoni*. Rock music in Slovenia became prevailing in subcultural events such as festivals, gigs and other gatherings. The popular Slovenian rock and punk bands include: *Bele Vrane*, *Buldožer*, *Lačni Franc*, *Lajbah*, *Pankrti*, *Videoseks* and many others, while the most famous solo performers include Tomaž Domicelj, Marjeta Neca Falk, Andrej Šifrer and others.

2.4 YU Rock Scene in Bosnia and Herzegovina

In Yugoslavia, Bosnia and Herzegovina was a fertile soil for the development of pop and rock music, which was especially evident in Sarajevo, Travnik, Bugojno and Bihać. One of the first rock groups in these territories was the group *Indeksi*, to which Kornelije Kovač joined later. After *Indeksi*, numerous popular groups emerged on the rock scene of Bosnia and Herzegovina, including: *Pro arte*, *Teška industrija*, *Ambasadori*, *Kodeksi*, *Bijelo dugme*, *Formula 4*, *Kongres*, *Divlje jagode*, *Vatreni poljubac*, *Zabranjeno pušenje*, *Bombaj štampa*, *Elvis J. Kurtović & His Meteors*, *SCH*, *Valentino*, *Plavi orkestar*, *Merlin*, *Hari Mata Hari*, *Crvena Jabuka* and many others, as well as solo performers who were previously or subsequently members of the listed groups: Seid Memić Vajta, Zdravko Čolić, Ismeta Dervoz-Krvavac, Jadranka Stojaković, Haris Verešanović, Željko Bebek, Goran Bregović, Mladen Vojičić

Tifa, Alen Islamović, Elvis Dž. Kurtović, Dino Merlin, etc. Želimir Altarac Čičak should be particularly singled out as one of the most significant and most influential creators of Sarajevo music scene, the editor and host of numerous popular radio and TV shows, as well as organizer of many music events, who paved the way to success for many unrecognized rock bands. Many of the said groups and performers became very appreciated in the world. Although the disintegration of Yugoslavia mostly influenced the future of music groups from Bosnia and Herzegovina, music which was created then is still listened by all generations in this territory.

2.5 YU Rock Scene in Macedonia

In Macedonia, particularly in Skopje, rock music began to be established in the 1960s. *Bezimeni* and *Biseri* were among the first groups, and later they joined to form *Bis-Bez* and started to introduce ethno sounds. Their originality later influenced the compositions of the most famous Macedonian band *Leb i sol* led by the world famous and very reputable musician Vlatko Stefanovski. Apart from those mentioned, many other groups appeared on the Macedonian rock scene, the most famous being: *Aleksandar Makedonski*, *Anastasija*, *Arhangel*, *Bastion*, *Den za den*, *Mizar* and others. Radio Skopje significantly contributed to the development of rock music in Macedonia and breakthrough of young and unrecognized rock bands by broadcasting music from the territory of the Former Yugoslavia and the world-famous rock hits of the time.

2.6 YU Rock Scene in Montenegro

Out of all countries of the former Yugoslavia, rock music was least developed in Montenegro. The reasons were numerous, but the most important ones include cultural and material aspects, and the fact that it was difficult at the time to purchase good musical instruments at affordable prices. Still, there were tendencies for rock music as a musical genre to take roots in the territory of Montenegro, especially in Podgorica. Some movies deal with these issues, such as Momir Matović's *To ludo srce*, and rock bands *Entuzijasti*, *Stele*, *Džeferdari*, *Lordovi*, *Svjetla u tami*, *Mrtve duše*, *Strašne sjenke*, *Podgoričani*, *Noćni putnici*, etc. These bands did not make great success, since only some of them managed to make several demo recordings, and they are little known today. Thus they were not able to succeed on their own, let alone some larger music scene. The greatest rock musician of Montenegro is Miladin Šobić, who was actively involved with music in the early 1980s. Although they did not have their own music scene, the youth was greatly influenced by rock music from the other territories of the former Yugoslavia.

3 Influence of Rock and Roll on Art, Media and Social Events

Influence of Rock and Roll was visible very fast in many social spheres. Presence of rock music in media influenced its increasing popularity. Rock and roll reached wide audience through radio and TV shows, magazines, books and movies. Soon, festivals devoted to rock music began to be organized, as well as gigs in increasing number of discotheques across Yugoslavia. First record labels began to be established, and the rock music was followed by numerous controversial phenomena. In this study, we have decided to list the most significant aspects of rock and roll on the media and social events, by dividing them into the following topics, which simultaneously correspond to the division of the project itself: Gigs and Discotheques; Festivals; Poetry and Books; Albums and Record Labels; Cult Radio Shows; Cult TV Shows; Filmography; Controversies and Magazines.

3.1 Gigs and Discotheques

Since 1977 bands like *Limunovo drvo*, *Igra staklenih perli* and *Električni orgazam* appeared in the *Student Cultural Center* in Belgrade, thus slowly forming the starting point of experimental theatre and alternative rock – the "New Wave". As the said "New Wave" was expanding, an increasing number of bands performed in the *Student Cultural Center*, among following bands stood out: *Pankrti*, *Šarlo Akrobata*, *Idoli*, *Urbana gerila*. The period of commercialization followed, along with ticketing and opening of an increasing number of discotheques across Yugoslavia.

In Belgrade alone there were several discotheques. *Cepelin* was usually frequented by young people, leaning towards the hippy philosophy, while *Akvarijus* gathered somewhat more serious members of rock population.

3.2 Festivals

As a genre, rock and roll developed from its modest beginnings of young bands in Yugoslavia which participated in the first gigs and dance parties, to first concerts and festivals. The period of rock festivals in Yugoslavia started in 1961 with *Omladinski festival* 'Youth Festival' in Subotica. Many festivals had campaigns with leaflets and posters. Rock festivals were primarily meant to connect people, promote rock and spread its influence. Some of the most significant rock festivals that were organized in Yugoslavia include the following: *Parada ritma* 'Parade of Rhythm', *Vatromet ritma* 'Fireworks of Rhythm', *I festival beat muzike* 'I Festival of Beat music', *I festival VIS-ova* 'I Festival of Vocal and Instrumental Ensembles', *II festival VIS-ova* 'II Festival of Vocal and Instrumental Ensembles', *I Šampionat VIS-ova Jugoslavije*

'I Championship of Yugoslav Vocal and Instrumental Ensembles', *I Beogradska gitarijada* 'I Belgrade Guitar Fest', *II Beogradska gitarijada* 'II Belgrade Guitar Fest', *I Sarajevska parada* 'I Sarajevo Parade', *Takmičenje beogradskih i sarajevskih VIS-ova* 'Competition of Ensembles from Belgrade and Sarajevo', *I Jugoslovenski festival Beat muzike* 'I Yugoslav Fest of Beat Music', *I Festival jugoslovenske pop muzike* 'I Festival of Yugoslav Pop Music', *II festival jugoslovenske pop muzike* 'II Festival of Yugoslav Pop Music', *Boom festival* 'Boom Festival', *Gitarijada* 'Guitar Fest' in Zaječar, *Omladinski festival* 'Youth Festival' in Subotica and *FAMUS* 'Festival of Acoustic Music in Sivac'.

3.3 Poetry and Books

Rock music also found its place in books and poetry. The Project Multimedia Document 2014/15 lists and summarizes approximately 70 publications related to the rock scene in Yugoslavia. Publications such as *Ilustrovana YU rock enciklopedija 1960–2006* written by Petar Ignjatović (Janjatović, 2001) and *Leksikon YU mitologije* (Andrić et al., 2012) describe development of rock music in Yugoslavia, provide biographies and photographs of performers, radio and TV shows, festivals, etc. The following books should be singled out: *Ljubav je samo reč: izbor ljubavne poezije jugoslovenske zabavne i rok muzike*, edited by Vojislav Nestorović (Несторовић, 2003) and *Obična ljubavna pesma* edited by Nenad Atanasković and Miroslav Josipović Atanasković and Josipović (2004). Books which are listed and described in this special topic were particularly helpful during the preparation of the project. Owing to the authors of the listed publications, an abundance of information concerning all aspects of Yugoslav rock scene is available to us.

3.4 Albums and Record Labels

First record labels in Yugoslavia emerged after World War II. Jugoton was the first record label in Yugoslavia and it was established in 1947. Production of gramophone records began in 1956 in Yugoslavia.

PGP RTB record label was established in Belgrade in 1951. The most important albums issued by this record label include *Pub* by Đorđe Balašević (1982) and *Ujed za dušu* by Riblja Čorba (1987).

Record label Diskoton was established in Sarajevo in 1973. The following groups issued albums for them: *Bijelo Dugme*, *Zabranjeno pušenje*, *Indeksi*, *Divlje jagode* and many others.

In 1972 in Zagreb, agency Suzy started operating, which did not have its own mechanism for producing records but it rather used services of other producers.

Among others things, the Agency organized concerts of *Parni valjak*, *Azra*, *Prljavo kazalište* and other famous performers.

In the book *YU 100: najbolji albumi jugoslovenske rock i pop muzike* 'Best albums of rocks and pop music in Yugoslavia' by Dušan Antonić and Danilo Štrbac (1998) (Kostić, 2013b), *Bijelo Dugme* and *Riblja Čorba* stand out, each with 8 albums. Record label Jugoton published the largest number of albums, 47.

3.5 Magazines

Magazines with rock topics quickly gained high popularity in Yugoslavia, especially in younger population. They reported about concerts, made posters and top lists of the most listened songs, and made very substantive interviews with performers. Some of the greatest stars about which the magazines wrote included *Riblja Čorba*, *Indeksi*, *Idoli*, while foreign stars, such as John Lennon, Bob Marley, and many others were also included. Some of the magazines published in Yugoslavia include the following: *Ritam*, *Džuboks*, *Ladin Džuboks*, *Vreme zabave*, *Ju rok magazin*, *Ukus nestašnih*, *X zabava*, *Pop Express*, *Heroina* and *Rok 82*.

3.6 Cult Radio Shows

The "New Wave" spread throughout Yugoslavia from Belgrade and Zagreb. Presence of rock music on radio was of key significance for the breakthrough of performers and presentation of their hits to as many people as possible. Persons who have greatly contributed to radio shows related to the YU rock scene include Zoran Modli, Nikola Karaklajić, Nikola Nešković, Vlada Džet, Nikola Kanjevac and many others. Radio Belgrade broadcasted popular shows, such as: *Zeleni megaherc*, *Veče uz radio*, *Prijatelji zvezda*, *Sastanak u 9 i 5*, *To je samo rokenrol*. Radio Belgrade 202 broadcasted *Ventilator* and *Hit 202* While *Diskomer* was broadcasted by Studio B. Radio Novi Sad broadcasted *Randevu sa muzikom*, while Radio Požarevac broadcasted *To je samo pop i rok*. Radio Luxembourg was a radio station that was very popular across Europe, as well as in Yugoslavia.

3.7 Cult TV Shows

Rock music gained its share on television after radio and magazines. Once Radio Belgrade became Radio Television Belgrade in 1958, the new television in Serbia, like in other states of Yugoslavia, was primarily formed by professions from radio, film, theatre and press. The programmes followed newspapers in creation of television content, as well as the development of technology. Drama series, documentaries

and entertainment programmes were also broadcasted, as well as educational programmes, children programmes and movies. When it comes to rock music, after occasional reviews in TV Dnevnik (TV Journal), the first show was made which dealt with the rock scene – *Koncert za ludi mladi svet*. Wider audience was thus allowed to get informed about concerts, new hits and albums of an increasing number of performers. The following are considered the most important cult TV shows, which regularly or occasionally reported on everything related to Yugoslav rock scene: *Hit meseca*, *Stereovizija*, *Rokenroler*, *Formula I*, *Beograd noću*, *Petkom u 22*, *Koncert za ludi mladi svet*, *Rock oko*, *Top lista nadrealista*, *Maksimetar* i *Garaža*.

3.8 Filmography

Starting from the late 1950s, until the mid 1960s, music unobtrusively appeared in the movies *Subotom uveče* (1957) directed by Vladimir Popović, *Ljubav i moda* (1960) by Ljubomir Radičević and *Zvižduk u osam* (1962) by Sava Mrmak. In movies *Kad budem mrtav i beo* (1967) directed by Živojin Pavlović and *Nemirni* (1969) directed by Kokan Rakonjac, owing to groups such as *Elipse*, *Zlatni dečaci* and *Siluete*, rock music was more and more present. In the 1980s, mixture of rock music and movies reached its peak in the movies *Davitelj protiv davitelja* (1984) directed by Slobodan Šijan, *Dečko koji obećava* (1980) by Miloš Radivojević, *Crna Marija* (1986) by Milan Živković and *Šest dana juna* (1985) directed by Dinko Tucaković. Kornelije Kovač, Vojkan Borisavljević, Zoran Simjanović and Vlatko Stefanovski are among the musicians who particularly stand out for their contribution to the unification of rock music and movies. Many musicians starred in movies, while actors became musicians.

3.9 Controversies

Some musicians found a way to express their opinions and attitudes, and stand out on the music scene, using lyrics which were then considered controversial, album covers, clothes or make-up. Public was thus, among other things, shaken by songs such as *Kurvini sinovi* (*Azra*), *Crni mercedes*, *Na zapadu ništa novo* (*Riblja Čorba*) and *Kralj alkohol* (*Time*). Great attention was stirred by the album cover of Riblja Čorba – *Pokvarena mašta i prljave strasti*. Long hair, unusual make-up, lyrics that openly indicated problems in society or topics which were considered taboo, all of it was in a certain way considered scandalous.

4 Gathering, Processing and Shaping of the Contents

For this project students acquired articles mainly from encyclopedias, newspaper and periodicals libraries and primary and secondary monographic sources. Rest

of the material – photos, audio tracks and videos were found either in cultural and media institutions (libraries, archives, radio stations, television networks, movie libraries and studios) or on the World Wide Web. In four months, over twenty institutions in seven cities had been visited and cooperation was established with many individuals who helped the project develop. The research was conducted by students divided into groups.

Several groups researched the rock scene in parts of former Yugoslavia: Belgrade (Aleksandra Arsenijević, Milica Ninković and Milena Obradović), rest of Serbia (Aleksandra Kojić), Croatia (Jovana Došenović, Maja Ivančić and Marko Petrović), Bosnia and Herzegovina (Violeta Kolaković, Milica Perišić and Petar Popović), Slovenia (Mihailo Škorić), Macedonia and Montenegro (Aleksandar Pavlović).

Rest of the groups researched influence of rock and roll on art, media and social events: gigs and discotheques (Katarina Jugov, Ivana Merdović and Jelena Radić), festivals (Milan Avdulah and Uroš Tošić), books and poetry (Marija Brašanac and Tamara Stanojčić), albums and record labels (Snježana Mirčeta and Ivan Peić), magazines (Maja Miladinović and Katarina Radovanović), cult radio shows (Milica Đorđević), cult TV shows (Ivana Merdović and Milena Obradović), filmography (Aleksandra Arsenijević and Jelena Radić) and controversies (Vladana Nešović and Uroš Timić).

After the gathering of information and material was completed, it was supposed to be processed, systematized and displayed in a desired manner. Mihailo Škorić and Petar Popović got that job.

Processing of materials related to territorial division got on the table first. Members of the student groups who researched it got an assignment to arrange the data in the given way, so that all material can be processed together and in the same way. Decimal classification was used – every piece of data was given an identification number. First digit indicated the region, second digit indicated a city from that region, third and forth digit indicated a music band or artist from that city, and fifth and sixth indicated an album published by the band. All data together with given numbers was put into an Excel spreadsheet, and then transformed into a SQL database.

Table 1 illustrates the way data is stored inside the database. Column GID (group identifier) is combined identifier of a music group (first digit – 1 – indicates that group is from Serbia, second digit – 1 – indicates it is from Belgrade and last two are unique identifiers so – 01 – first group from Belgrade, Disciplina kičme). Other columns are MID (city identifier – 11 – Belgrade, Serbia), and columns with names of the files for: figure of the band, its biography, discography and a fragment of their song. Data redundancy is obvious (repetition of identifier parts), but it was necessary to increase the control and rate of finding possible mistakes, which were inevitable because the base was filled by many different people. After the data

was harmonized with the database, all was transferred to a local server for further analyzing and testing of the website so that it would display the acquired data in a desired manner.

GID	Group	MID	Photo	Bibliography	Discography	Song
1101	Disciplina kičme	11	1101.jpg	1101b.txt	1101d.txt	1101.mp3
1102	Divlji anđeli	11	1102.jpg	1102b.txt	1102d.txt	1102.mp3
1103	Dogovor iz 1804.	11	1103.jpg	1103b.txt	1103d.txt	1103.mp3

Table 1. Example of a part of the *grupe* table

Figure 1. Example of a region page – Slovenia

First the basic composition was made with hierarchy architecture in its core. Instead of making HTML pages for each and every region, city or band, four PHP-scripts that consult the SQL-database were made and depending on the user's choice dynamically generate proper HTML content.

First script, *mapa.php*, shows the map of Yugoslavia with proper borders and capitals. Sections of the map are also links to *rep.php* script along with proper parameter forwarding.

The *rep.php* script shows the content for one specific region (or republic), generates the list of its cities and their figures. figure of every city is a direct link to the page of the selected city (Figure 1).Under the city list *rep.php* generates a text about the rock scene in the region, while above the list it generates the links to all other regions.

Figure 2. Example of a city page – Ljubljana

By activating the figure hyperlink, user is redirected to *grad.php* script which generates the list of performers from the specified city along with their respective figures. On the top of the page there is navigation (links to other cities of the same region) (Figure 2). In the background the figure of the city is generated. Every performer's figure is a hyperlink to *bend.php* script along with the proper parameter forwarding (Figure 3).

The last script, *bend.php* has only one link – to go backwards. Role of this script is to generate the performer's biography and discography. For every album figure of the cover is generated, along with the name of the album, year of debut and its publisher company. After the page is loaded, a tune is automatically reproduced, with a large photo in the background.

Figure 3. Example of a performer's page – Buldožer

As there are many possible outcomes of the pages content, it is necessary to find them an appropriate theme (color) to share. Main color was chosen to be a shade of red (`\#dd2233`), with strawberry themed heading along with the serif font *prociono*. Heading which all pages share consists of a main menu, audio player and search bar. They are let floating using the CSS, thus sacrificing the flexibility for looks, on the assumption that the site would be viewed from a standard resolution display.

Figure 4. Home page display

On the same assumption, home page was made to show a room (Figure 44) which, aside the map with regions, contains other objects as links to redirect the user to pages of other specialized sections that make the music scene: festivals, gigs, discos, magazines, filmography, radio and TV shows, record publishers, books and controversies which the Yugoslavian rock scene made and influenced.

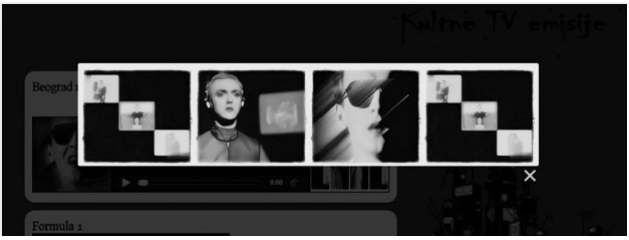


Figure 5. Example of photo viewing inside the gallery

Pages about social aspects of rock and roll contain articles written by students, as well as multimedia content including figures, video and audio tracks. Background

is green (`\#E1F5A9`), usually together with background figure. Material is displayed using *JavaScript* that runs when you click the specific photo or track. All figures can be skimmed trough and opened figure tabs can be closed by clicking (`x`) button (Figure 5). Videos or audios are played via integrated players (Figure 6).

As mentioned before, each page header contains a search bar that can search trough the database (names of places, performers and albums). If the user doesn't know where the performer is from but knows his name, he can type it into the search box and find information more easily. Based on the entered keywords a query is formed and the wanted data is displayed if found (Figure 7).

Figure 6. Example of video viewing trough player

5 Students' Experience of the Projects – Positive and Negative Aspects

Work on the project entitled YU ROCK SCENE, within the subject Multimedia documents 2014/2015, was invaluable for all senior students studying Library and Information Science. We had the opportunity to unify and apply the knowledge gained during our four-year study programme, but we were also able to gain new knowledge. The topic of the project is very interesting and close to us. We have successfully cooperated with significant institutions and individuals. Project Multimedia Document 2014/2015 enabled all of us to express our interests and skills.

On the other hand, the downside was the extent of the topic. Selection and organization of information represented a great challenge. We have invested effort to select the most important information and represent it in an understandable way. Good organization and team work skills were necessary since many students participated in this project.

Multimedia document may be accessed on the website of the Faculty of Philology at the following address: `http://www.fil.bg.ac.rs/mmd_27/mmd_2015/home.php`.

```
<?php
$query = "SELECT Naziv, Ime, grupe.GID, Godina, Omot FROM albumi, grupe
WHERE albumi.GID = grupe.GID AND INSTR(Naziv, '{$grad}') > 0";
    $result = mysql_query($query, $link);
        if($result === FALSE)
        {
        die("Upit nije vratio ništa");
        };
        $count = 0;
            while ($row = mysql_fetch_array($result, MYSQL_BOTH))
                {if($count==3)
                    {   echo "<tr>";
                        $count = 0;
                    }
                if($count==0)
                    {   echo "<tr>";
                        echo "<td>";
                    }

echo "<p class='dgra'>" . $row['Naziv'] .  ", </p></a>";
echo "<p class='dgra'>" . $row['Godina'] .  "</p></a>";
echo "<a href='bend.php?x=" . $row['GID'] .  "' class='dgra'><p>" . $row['Ime'] .  "</p></a>";
$adresa = "OMOTI/" . $row["Omot"];?>
```

Figure 7. Script that searches database for keywords and displays found results

6 Acknowledgements

We owe great gratitude for the development of this project to professor Cvetana Krstev, the project holder, and to professor Miloš Utvić and coordinator Biljana Lazić. Their assistance, guidance and ideas were of paramount importance. We would like to thank the following individuals and institutions for their support and cooperation, for allowing us to use the rich information resources and for sharing the significant information on the Yugoslav rock scene: Petar Ignjatović, Željko Radišić, Miroljub Stojanović, Ivan Velisavljević, Borislav Stanojević, Marko Maršićević, Nada Petronijević-Ćović, Jovan Jovanić, Ljubinka Gavran, Edo Gracin, Branko Vurušić, Radoman Kanjevac, Zoran Modija, Srđan Nikolić, Sloba Konjović, Ljiljana Krneta, Vojin Šarčević, Darko Kocjan, Marko Savić, Jelena Stevanović, Slobodan Boba Novaković, as well as the following institutions: Radio Television of Serbia, Film Center of Serbia, National Library of Serbia, Belgrade City Library, Radio Novi Sad, Radio Belgrade, Radio 202, Radio Požarevac, Croatian Radiotelevision, Yugoslav Cinematheque, National and University Library in Zagreb, National and University Library of Slovenia (Narodna in univerzitetna knjižnica) and Student Cultural Center.

References

Andrić, Iris, Vladimir Arsenijević, and Đorđe Matić, ed. *Leksikon YU mitologije.* Beograd : Rende, Zagreb : Postscriptum, 2012.

Atanasković, Nenad and Miroslav Josipović, ed. *Obična ljubavna pesma : antologija YU Rock ljubavne poezije.* Beograd : Mono & Mañana, 2004.

Fajfić, Željko and Milan Nenad. *Istorija Yu rock muzike.* Laćarak : Tabernakl, 2009.

Janjatović, Petar . *Ilustrovana YU rock enciklopedija : 1960–2000.*. Beograd : Prometej, 2001.

Kostić, Petar. „Šta su nama kič i šund?". `http://balkanrock.com/`, link `sta-su-nama-kic-i-sund` приступљено 20.03.2016, 2013a.

Kostić, Petar. „Yubilej: YU 100 – najbolji albumi jugoslovenske rok i pop muzike (15 godina kasnije)". `http://balkanrock.com/`, link `yubilej-yu-100-najbolji-albumi-jugoslovenske-rok-i-pop-muzike-15-godina-kasnije` приступљено 20. III 2016, 2013b.

Несторовић, Војислав. *Љубав је само реч: избор љубавне поезије југословенске забавне и рок музике*. Београд : Београдска књига, 2003.

Social Sciences and Computing: Master study program review

PAPER SUBMITTED: 03 September 2015
PAPER ACCEPTED: 05 September 2015

Vladan Devedžić
devedzic@gmail.com
University of Belgrade
Faculty of Organisational Sciences

Milan Krstić
milan.krstic.fpn@gmail.com
University of Belgrade
Faculty of Political Sciences

The graduate study program Social Sciences and Computing at the University of Belgrade is developed as part of the Tempus project INCOMING (Interdisciplinary Curricula in Computing to Meet Labor Market Needs, project No. 530155-TEMPUS-1-2012-1-EE-TEMPUS-JPCR). The program is designed as a continuation of undergraduate studies in a field of the broad study area of social sciences (economics, sociology, psychology, law, finance, management, languages, and so on, or a combination of some of these fields), but with a focus on application of modern computer technology in these fields.

Studying in this study program assumes that the students have already acquired some knowledge in a subset of these social sciences. It is also understood that students already have a basic knowledge of computer technology and that they are familiar with the basics of using computers and the Internet. Through the graduate study program Social Sciences and Computing, students are introduced to some of the more advanced techniques of using computers and the Internet, but always through application in a selected field of social sciences. A wide variety of courses offered allows each student to focus on the social science(s) that she/he is most interested in.

1 The program at a glance

It is a 60 ECTS, 1-year program, structured in such a way that students take courses in the Spring semester (typically 5 courses, 30 ECTS altogether), and then work on their term paper (qualification paper, 10 ECTS) and master's thesis (20

ECTS). Graduates from this study program receive the title Master of Computing in Social Sciences.

The program is modeled after a number of similar programs from EU and North American universities. Thus it has a strong flavor of quantitative disciplines and statistics, interwoven with numerous topics from the broad fields of social sciences and computing. There are two required courses:

- Contemporary computer technologies;
- Quantitative modeling in social sciences.

They are prerequisites for a wide spectrum of elective courses:

- Social network analysis;
- Data analysis and visualization;
- Demography and new information technologies;
- Research methodology and statistics;
- Introduction to cognitive linguistics;
- Digital humanities;
- Digitalization and transdisciplinarity in humanities;
- Programming for linguists;
- Legal and ethical aspects of ICT;
- Cybercrime;
- Quantitative methods in economics;
- Prediction and decision methods;
- ICT and sustainable development;
- Measuring information society;
- Security investment analysis;
- Technology enhanced learning;
- Computer science and music;
- Digital libraries;
- Application of information theory on language processing;
- Social psychology of cyber space;
- Applications of sampling theory in social sciences;
- Regulation of content and freedom of expression on the internet;
- Human resource development and ICT;
- Knowledge management and ICT;
- Risk management in actuarial science;
- Application of information systems in financial mathematics;
- Digitalization and traductology.

There are no formal restrictions in choosing electives, but the students are expected to demonstrate some proficiency in the broader field related to an elective they choose.

The program was launched in February 2015. Although at the time of writing this review (Early September 2015) no student has graduated from the program yet, all the courses have been already given to the first generation of students and most of the students have already passed the exams. Hence it was possible to do evaluations of the program (at the end of the Spring semester) and generate related statistics.

2 Teachers' view

One thing that most of the teachers who participated in the program in Spring 2015 (about 50 teachers altogether) have emphasized as a strong upside of the program was motivation of the students who have attended the classes. The teachers were happy to have good students, eager to learn new things, hardworking and very active during the classes. It is a common impression of most teachers that the students were capable of grasping the new topics very quickly. This is probably no wonder at all, since the students enrolled in the first generation (36 students) have come with extremely good GPAs from their BA programs. Given the fact that many students did not have previous knowledge in statistics and quantitative disciplines, and that their knowledge of computing was in many cases limited to standard office programs and Internet browsers, all the courses given went very smoothly, the teachers say.

An obvious downside was the heterogeneity of the students' previous knowledge, which has necessitated splitting the students in some of the classes in two groups (absolute beginners and those with minor experience with the course topics). Luckily, this has been compensated by enthusiasm from both the teachers' and the students' side, as well as by the fact that in most courses there were at least two teachers.

3 Students' view

In the online evaluation form that the students were asked to fill out in the end of the first semester, there were questions about:

– the quality of the courses given (the knowledge acquired and how useful the students perceive it, the compliance with the announced course syllabi, the perceived compliance with similar master programs in well-ranked universities worldwide, the use of modern teaching approaches and tools, the teaching materials and resources (literature, software, ...) available to students, online or in hard copy);

– the perceived relevance of the degree received by completing this study program;
– the perceived quality of the teaching and pedagogy demonstrated by the teachers;
– how well is the study program adapted to the level of mastery of computing typical for students who have majored in social sciences;
– the quality of the facilities (teaching rooms, equipment, Internet, ambient, ...);
– the demonstrated level of responsiveness and cooperation demonstrated by teachers and administrative staff alike.

20 students (out of 36) have submitted their evaluations. In all questions, they were supposed to answer using a 5-point Likert scale (5 being the highest grade). They were also allowed to write free-form comments.

It felt good to find out that the students liked the program in their evaluations. The mean values obtained in all answers were always higher than 2.5, and often higher than 3 as well. Figure 1 illustrates some of the evaluation results.

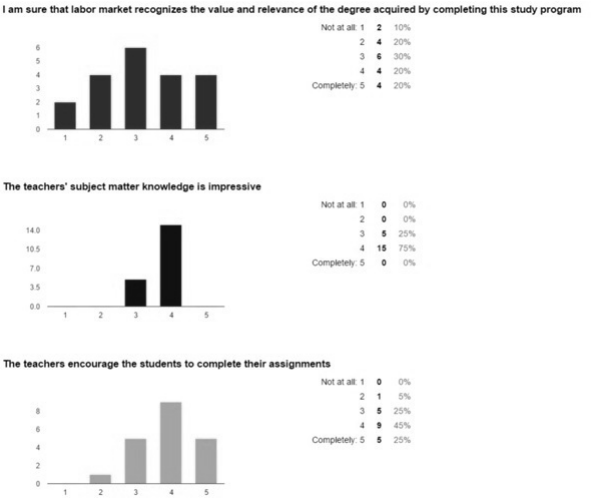


Figure 1. Some of the students' responses

The students have highly appreciated the fact that the program offered the possibility for gaining many practical computing skills, and saw it as a distinctive feature of the program. Therefore, this program appeared to be a suitable "plug in" to the broad theoretical and practical knowledge in different fields of social sciences that the students have received their BA degrees from. The opportunity to learn how to

use many software applications in various fields of their interest was the most important outcome. The students were generally very satisfied with the demonstrated knowledge of the teachers and mostly satisfied with their encouragement to complete the assignments. These two factors gave important contribution to the success of the program. Additional positive incentive was the fact that group included very active, highly motivated and competitive fellow students.

In addition, it is important to emphasize that the students were very satisfied with the variety of the elective courses offered to them. This fact allowed them to adapt the course selection to their interests and use the possibility to acquire new knowledge and skills in the fields relevant for their future career. Basic computing skills are today considered a necessity in the labor market, whereas more advanced skills offered by this program might be one of the potential comparative advantages to the other degree holders in social sciences.

An obstacle for some students might have been the very intensive pace of the program, especially during the first semester. In order to maximize the possibility for gaining new skills and knowledge, it was of vast importance to complete all of the assignments and regularly visit the lectures that have been held very frequently. Therefore, only highly motivated students should consider enrolling into this exceptionally interesting and useful graduate program.

4 Lessons learned

All in all, the evaluation results and the personal experiences of both the students and the teachers after the first semester are rather positive, but there is still room for improvement. For example, it has turned out from the evaluation forms submitted that the students were not quite happy with the compliance of the topics covered in classes with those announced in the course syllabi. It indicates that the teachers should better explain their wish to present the most current topics and the most current methodologies and tools, rather than sticking to the course syllabi strictly. Likewise, not all students have thought that the labor market will readily recognize the value and relevance of the degree acquired by completing this study program. Although the program creators have tightly collaborated with the Chamber of Commerce and associations of employers in Serbia when designing the courses, taking into account the labor market needs, it certainly requires more effort to make the students fully recognize these advantages.

Somewhat higher level of flexibility is also expected from the teachers in terms of the assignments – some of the students thought that they were asked to do assignments that require a rather high level of prerequisite technical knowledge (which few of the graduates from BA programs in social sciences have).

5 To be continued

Most of the first-generation students enrolled in the study program Social Studies and Computing are expected to graduate by February 2016, when the new generation of students will start with the classes. It will be interesting to make a survey of their theses' topics at that time (many of them are currently in the process of picking the theses' topics), in order to better understand the students' interests and expectations.

Meanwhile, the positive atmosphere around this study program continues. Potential applicants keep contacting the program council members about details, and it is encouraging to see that they come from a variety of BA study programs in social sciences, as well as from different universities. As an illustration, here's a comment from a potential applicant (holding a BA degree in Philosophy) sent to the Program Chair along with an inquiry about the program:

I think that nowadays it really makes sense to enroll in such a master's program, since in high-ranked universities like Stanford students of Philosophy take courses in Philosophy together with courses in Computer Science, Linguistics, and Psychology. It is the only way to really ***apply*** *Philosophy.*

Review of the 2015 EUROLAN in Computational Linguistics

Jelena Mitrović
jmitrovic@gmail.com,
University of Belgrade
Faculty of Philology

PAPER SUBMITTED: 10 February 2016
PAPER ACCEPTED: 14 February 2016

EUROLAN Summer School (13–25th July, 2015) was the twelfth in the series of summer schools that take place every two years in Romania. Topics of these summer schools are always very "hot", as was the case with this year's topic – an area of Computational Linguistics and data and knowledge management in general – Linguistic Linked Open Data – LLOD. This year's summer school took place in the beautiful town of Sibiu, situated in the heart of Transylvania, surrounded by mountains and wonderful nature.

Twenty distinguished lecturers who are some of the biggest names in the area of Computational Linguistics, and certainly in the LLOD movement, and who significantly contributed to its development, held intensive courses during two weeks of the summer school. Mornings were mainly reserved for theoretical parts of acquiring new concepts, while the afternoons were filled with practical work, tutorials and practice. Participants came from all over Europe, but also from China and Australia. One of the benefits of this kind of learning and enhancement of professional skills is also the possibility to share experiences and to connect with colleagues from all over the world – which is definitely very important! This was the first summer school I attended and it was a great pleasure to spend two weeks with like-minded people who share my passion for linguistics and using computer technologies to process natural language.

The concept of Semantic Web was in the centre of all lectures and tutorials at this summer school. Semantic Web is a project of developing a universal medium for exchange of information using documents with meaning that computers can process on the web. The main goal of this concept is semantic interoperability of web resources and existence of an infrastructure for machine interpretation and reasoning about the contents on the web.

RDF (Resource Description Framework) is a concept that originated from the search for efficient solutions in information retrieval tasks and is one of the standards of Semantic Web. It represents a general method for conceptual description of information – semantic links between electronic resources. It consists of ordered

triplets: Subject – Predicate – Object, where the Subject is the RDF URI reference of the resource we are describing, the Predicate is the RDF URI reference, i.e. the semantic relation, and the Object is the RDF URI reference, the meta-datum itself.

Key technology of the Semantic Web is also SPARQL (pronounced as the word "sparkle", a recursive acronym for SPARQL Protocol and RDF Query Language). This language was developed specifically for searching RDF databases and is a W3C standard. Other important elements of Semantic Web are XML (eXtensible Markup Language), which defines the data structure (RDF/XML), ontologies, i.e. models that represent knowledge or sets of concept definitions and relations that exist between them, OWL (Web Ontology Language) which is used for ontology publishing and sharing. All of these technologies are crucial for functioning of the LLOD paradigm.

Linked Open Data – LOD, the basis for LLOD, according to the principles set in 2006 by Tim Berners-Lee, are data that 1) Use URIs (Uniform Resource Identifiers) as names of things; 2) Use HTTP URIs links so that those names can be found 3) Provide useful information using RDF and SPARQL standards; 4) Contain links to other URIs for discovering as many things as possible; 5) Data should be open for usage via open licences.

Graphical representation of the LLOD cloud was developed following the initiative of the Open Linguistics Working Group [1] and this group is responsible for its development and maintenance, in the scope of the Open Knowledge Foundation Network (OKFN) [2]. Figure 1 shows the LLOD cloud diagram containing corpora, databases, terminological bases, dictionaries, linguistic data categories, typological databases.

DBpedia is a very important part of both LLOD and LOD paradigms. It transforms data from Wikipedia pages into RDF. It contains URIs and other metadata for each page, starting with infobox parts of Wikipedia pages. BabelNet is also a very important element of the LLOD cloud. It is a semantic network that automatically extracts data from WordNet, Open Multilingual WordNet (set of all open wordnets), Wikipedia (the largest collaborative encyclopaedia), Wikidata (the largest collaborative knowledge base), Wiktionary (the largest collaborative dictionary), OmegaWiki (medium-sized multilingual dictionary).

A lexical resource can be included in the LLOD cloud if the following requirements are fulfilled: 1) the resource has to be available through resolvable http:// (or https://) URI links; 2) the data have to be resolvable into RDF data in one of the mostly used RDF formats (RDFa, RDF/XML, Turtle, N-Triples); 3) It has to contain at least 1000 ordered triples; 4) It has to be connected through RDF links

[1] OWLG `http://linguistics.okfn.org/2011/05/20/the-open-linguistics-working-group/`

[2] OKFN `https://okfn.org/`

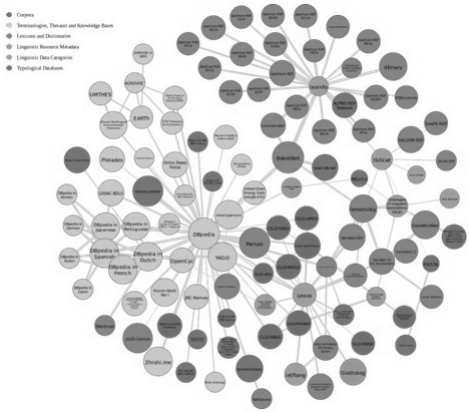


Figure 1. LLOD cloud

with a resource that is already included in the LLOD diagram, or it has to have at least 50 links to other resources; 5) It can be accessed through RDF crawling, through an RDF dump, or through a SPARQL endpoint.

At this summer school, I also had a chance to find out that the idea of accepting the LLOD technologies has been accepted in the WordNet community, that is to say, LLOD as the basic mechanism for creating links between wordnets in different languages, through the interlingual index (ILI). Accepting open licences and shared formats has led to more available data from world's wordnets. One of the projects that stemmed from that notion is the Universal wordnet whose goal is to resolve the problems of polysemy and synonymy through connecting the same or similar concepts in different languages.

The conclusion of the story about Linguistic Linked Open Data is that they are a useful solution for many purposes because they: 1) allow for integration of information – it is possible to find and combine information from various resources in an efficient way; 2) enable multilingual usage for many tools; 3) enforce dynamic publishing – data on the web are not static, different versions can be seen and errors corrected; 4) use graph based models which allow for representing any kind of a linguistic resource; 5) information retrieval is structured, e.g. we can get an answer to the question "What are the names of all Nobel Prize winners originating from France?".

Besides gaining invaluable knowledge about the way in which many Semantic Web technologies function and witnessing examples of their usage in hands-on sessions, I also received a lot of useful suggestions as to how we could include Serbian

linguistic resources and tools into the LLOD cloud, and all of that from experts in this field, some of which have invented or significantly enhanced the technologies we used. The EUROLAN School left a very positive impression on me, especially because the organizers insisted that all participants should spend as much time as possible together, which is why we had a chance to better acquaint ourselves with the lecturers and to get valuable advice. The next one in the series of these summer schools will be organized in 2017 and everyone who is interested in Computational Linguistics should attend it, maybe it will even be me again!

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 250 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 LaTeX(the journal style and all packages can be downloaded from the journal web site). Authors that are not familiar with LaTeXcan 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 LaTeXwill 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. itle of the paper should not be written in capital letters. The authors should keep the length of titles reasonable – preferably less than 150 characters. For titles longer than 90 characters 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 LaTeXin 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 LaTeXor 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).