

# Conference Program

**Monday, June 20<sup>th</sup>**

**9.00 Opening ceremony(Ceremonial Hall -Rectorat)**

**Prof. Dr Miladin Kostić, rector SUNP**  
**Prof. Dr C. Dolićanin, Academician of JANN**  
**Academician SASA T. M. Atanacković**

**COCKTAIL**

**9.15 Plenary Session (Ceremonial Hall -Rectorat)**  
**Chairmen**

**Academician SASA, S. Pilipović**  
**Prof. Dr C. Dolićanin, Academician of JANN**

**9.15-9.40 I. Gutman, Graph Energy - Old and new Results, Serbia.**

**9.45-10.10 T. Atanackovic, Viscoelasticity with fractional derivatives of real and complex order. Serbia.**

**10.15-10.40 M. Spies, Topic Modelling -- Effects of Deep Vocabulary Preprocessing. Germany.**

**Ceremonial Hall -Rectorat**

**10.45 Session 1**  
**Chairmen**

**Academician, I. A. Soloviev**  
**Academician SASA, M. Matijević**

**10.45-10.55 S. Pilipovic, G-type Spaces of Ultra distributions over  $R_+^d$ , Serbia.**

**11.00-11.10 E. I. Milovanović, Č. B. Dolićanin, I. Ž. Milovanović, Remark on the normalized Laplacian ratio spread of graphs. Serbia.**

**11.15-11.25 B. Mirić, P. Spalević, A. Pavlović, Characteristics of inductive loops applied in thelematics. Serbia**

**11.30-11.40 I. Ž. Milovanović, E. I. Milovanović, E. C. Dolićanin, Dž. Pičić, Remark on the forgotten index of graphs.**

**11.45-11.55 N. Grahovac, M. Žigic, Fractional derivative model of a frontal vehicle impact. Serbia.**

**12.00-12.10 B. Borovičanin, On the extremal Zagreb indices of trees with given parameters. Serbia.**

**12.15-12.25 E. Glogić, E. Zogić, N. Glišović, Remarks on the upper bound for Randić index of bipartite graphs. Serbia.**

**12.25 Lunch**

**13.30**

**Session 2**

**Chairmen**

**Prof. Valery Romanovski**  
**Prof. Dr Dragić Banković**

**13.30-13.40 V. Burégio, N. Faci, E. Kajan, Z. M. , Mohamed Sellami, Bringing semantics to the social web , France, Brazil, UEA, Serbia.**

**13.45-13.55 I. D. Arandelović, On semi-metric fixed point results, Serbia.**

**14.00-14.10 M. Petrović-Torgašev, A. Šebeković, Ricci pseudo-symmetries of generalised wintgen ideal legendrian submanifolds, Serbia.**

**14.15-14.25 A.M. Lykov, A.A Bashkatova, Calculation of the surface materials modification by plasma processes. Russia.**

**14.30-14.40 S. Maćešić, Ž. Čupić, L. Kolar-Anić, Method for detection of Andronov-Hopf bifurcation in the models of chemical reactions, Serbia.**

**14.45-14.55 Ž. Čupić, A. Ivanović-Šašić, S. Blagojević, S. Blagojević, S. Anić, L. Kolar-Anić, Slow manifolds of the model for Bray-Liebhafsky reaction analysed by return maps, Serbia.**

**15.00-15.10 U. Marovac, D. Banković, Applications of Boolean equations, Serbia.**

**15.15-15.25 N. Glišović, M. Rasković, The System for Decision Support in the Diagnosis of Systemic Autoimmune Diseases, Serbia.**

**15.30-15.40 A. Ljajić, A. Avdić, U. Marovac, E. Kajan, Creating domain dictionaries for serbian language, Serbia.**

**15.45-15.50 D. Pokrajac, M. Petkovic, Computational Geometry Issues In Recursive Partitioning Based Simulation, Algorithm, Serbia, USA.**

**15.50 Coffee break**

**Session 3**

**Chairmen**

**Academician SASA, T. M. Atanacković**  
**Prof. Dr D. C. Dolićanin-Đekić**

**16.10-16.20 T. M. Atanacković, G. Hörmann, M. Janev, Lj. Oparnica, S. Pilipović, D. Zorica, Generalization of the wave equation including hereditary and non-local effects, Serbia, Austria.**

**16.25-16.35 M Knežević, Some properties of hyperbolic partial derivatives for harmonic quasiconformal mappings, Serbia.**

**16.40-16.50**

**A. Dimov, L. Stoimenov, An approach for software reliability management in Big Data distributed stream computing systems, Bulgaria, Serbia.**

**16.55-17.05**

**E. Dolićanin, O. Kasum, A. Perović, A. Jovanović, A. Mumdzic Implementations, migrations and generalizations of some fundamental properties in computability models, Serbia**

**17.10-17.20**

**D. A. Romano, A note on regular coequality relation, BIH.**

**Conference Dinner**

**19.00**

**Tuesday, June 21<sup>th</sup>**

**Plenary Session (Ceremonial Hall -Rectorat)**  
**Chairmen**

**Prof. Dr I. Milovanović**  
**Prof. Dr M. Knežević**

**9.00**

**9.00-9.25**

**M. Ruzhansky, Schatten classes, nuclearity and nonharmonic analysis on compact manifolds with boundary, Russia, UK.**

**9.30 - 9.55**

**Valery Romanovski, Diana Dolićanin, n-tegrability and reversibility in polynomial systems of ODEs, Russia, Slovenia, Serbia.**

**10.00-10.25**

**E. Malkowsky, V Velčković, F. Ozger, Compact operators and visualizations, Germany, Turkey, Serbia .**

**(Ceremonial Hall -Rectorat)**

**10.25**

**Session 1**  
**Chairmen**

**Prof. Dr B. Mirić**  
**Prof. Dr Ejub Kajan**

**10.25-10.35**

**M. Matijević, M Albijanić, Convexity, inequality and other applications, Serbia**

**10.40-10.50**

**I. A. Soloviev, D. C. Dolićanin-Đekić, Stochastic models of heat transfer based on generalized equation of Fokker-Planck-Kolmogorov, Russia, Serbia**

**10.55-11.05**

**A. H. Ansari , S. Chandok, T. Došenović, S Radenović,  $\alpha\beta$ -contractive mappings and upclass of type II function, Iran, India, Serbia.**

**11.10-11.20** **T. M. Atanacković, M. Janev, S. Pilipović, D. Zorica,** Variational principles with Lagrangians containing derivatives of complex fractional order, Serbia

**11.25-11.35** **Ir. R. Visuer,** Success Factors of an Open Source Outsider: the Case of Nginx Web Server, Belgium.

**11.40-11.50** **R. Unland,** Industrial Agent Technology, Germany, UK.

**11.50** **Coffee break**

**12.05** **Closing of the Conference**

**12.25** **Sightseeing of the city and surrounding**

**Department of Mathematical Sciences**

organizes

**4<sup>th</sup> International Conference**

**CONTEMPORARY PROBLEMS OF MATHEMATICS,  
MECHANICS AND INFORMATICS (CPMMI 2016)**

**Conference Chairman**

**Prof. Dr Cemal Dolicanin, Academician of JANN**

State University of Novi Pazar

**Basic information**

**4<sup>th</sup> International Conference "CONTEMPORARY PROBLEMS OF MATHEMATICS, MECHANICS AND INFORMATICS "** will be held at The State University of Novi Pazar, **19, 20 and 21. June 2016,** Novi Pazar.

**Registration and Agenda**

**19. June 2016.** Registration at the State University of Novi Pazar, from 12:00.

**20. June 2016.** 9:00 Opening ceremony

**21. June 2016.** 9:00 Working sessions

12:05 Conference session ending

12:25 Sightseeing of the city and surrounding

**Conference Secretariat**

The correspondence should be performed via Conference Organizing Committee

State University of Novi Pazar

Vuka Karadzica bb

36300 Novi Pazar, Serbia

Tel:+381 20 317 754 ; +381 20 317 752

www.cpmmi.np.ac.rs ; e-mail:cpmmi@np.ac.rs

**Contact persons**

**Ahmed Mumdzic ,Damir Skarep, Maida Musovic,**

e-mail: cpmmi@np.ac.rs

**Accommodation**

**Accommodation** is organized in cooperation with State university of Novi Pazar, at the hotel "VRBAK". Hotel "Vrbak" is located in the center of Novi Pazar. Below the hotel, the river Raska flows. With its distinctive architecture, internal and external environment, it leaves a strong impression to the passers-by.



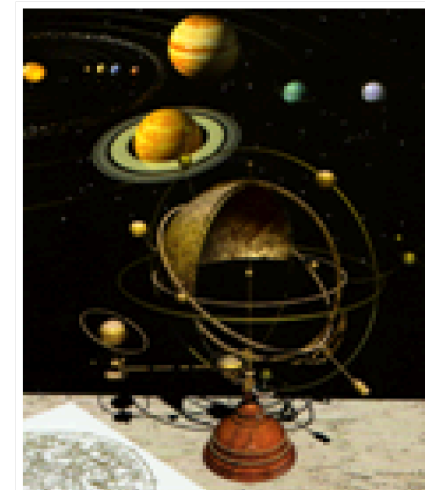
**State University of Novi Pazar**

State University of Novi Pazar



**Third Announcement  
CPMMI 2016**

**4<sup>th</sup> International Conference  
CONTEMPORARY PROBLEMS  
OF MATHEMATICS,  
MECHANICS AND  
INFORMATICS**



**Novi Pazar, 19-21 June 2016.**

# Creating Domain Dictionaries for Serbian Language

Adela Ljajić, Aldina Avdić, Ulfeta Marovac and Ejub Kajan  
*State University of Novi Pazar*  
*Vuka Karadžića bb, 36300 Novi Pazar, Serbia*  
*umarovac@np.ac.rs*

## Abstract

Automatically created thesauruses are used in order to improve methods for clustering, mining and determining the sentiments of some specific data corpus. There are different methods for the automatic discovering of similar words. Some of them are based on text corpora and mathematical similarity measures, while others use graphs and monolingual dictionaries. Serbian language is the richer than the English, by vocabulary and grammatical issues. Known methods for automatic thesaurus generation may neglect some of these specific issues. This paper deals with a method for automatic generation of a thesaurus from the repositories of documents in the Serbian language based on mathematical methods such as chi-square test, cosine similarity and Jaccard similarity coefficient. The proposed method can be applied either to normalized or non-normalized documents.

## 1 INTRODUCTION

With development of computer technology and context-aware computing, numerous data is collected on daily basis in order to obtain information that would facilitate user interaction with computer applications. An important task in this process is to determine which groups of documents carry similar information, and how to get these groups on automated way so that answers are as relevant and closer to human understanding, and yet more objective and more comprehensive than human conclusions in choosing: "What is similar?" .

Data mining is the process of finding interesting information in large collections of inadequately structured data [1, 2]. Interesting information is non-trivial, available, previously unknown and potentially useful information.

Data mining was created as a result of the growing need for market research and detecting potential customers and understanding and meeting the needs of existing ones. This is a method of data search, which developed rapidly thanks

to the development of computer technology. Only with the development of fast computer systems has made it possible to efficiently search large amount of accumulated data. Thanks to "data mining", data sets are converted into structured information and knowledge and thus are useful and valuable for research.

Text mining is a branch of data mining that seeks to find interesting patterns from large databases of text written in natural languages [3].

Thesauruses are specific dictionaries, lexicons that are typical for a particular domain. In the field of data analysis and text analysis, creating a thesaurus includes the formation of lists of synonyms, i.e. words that is related or similar to a given word in a given domain. A domain is defined by corpus of documents in which synonyms are looking for. Since that it is difficult to automatically distinguish between synonyms, antonyms and hypernyms, this term synonyms refers to the words with close semantic meaning.

Synonyms are words that have the same or similar semantic meaning, but differ in form, i.e. their grammatical composition is not the same. Therefore, the automated solution of this problem has long been a challenge, both in the field of artificial intelligence and in the field of text analysis. It is not only a challenge "to teach a computer to understand" that two apparently different words actually carry the same meaning, but after that computers objectivity should be used in the selection of what is similar, having in mind that every "human expert" could overlook, forget or neglect something, based on his own experience.

Depending on the type of corpus, different resulting list of synonyms could be obtained. Dictionary made from the corpus is specific to that corpus domain and is more adapted to a particular use in that field than the general dictionary made by people. There are several advantages for using computer-made thesaurus. In particular, they can be easily edited to reflect the changes in the collection of documents (in the appropriate field). These thesauruses are not based on the subjectivity of the writer (they are, of course, based on the corpus in use). It is obvious, however, that writing a dictionary of synonyms by the people should be more responsible, with less errors in total. According to two classical measures for finding information, it is expected that computer-made lexicons have a better response (or coverage) and a lower precision (except for the words whose meaning is much related to the scope of application) than general-purpose dictionaries of synonyms made by human [4].

This paper gives an overview of existing methods for generating thesaurus that have been used and applied for English language. They are divided into two groups, those that use ready-made lexical resources and those that are based only on the corpus of documents and mathematical methods. Serbian language is grammatically and linguistically much richer than the English language, with richer vocabulary and grammatical categories, such as cases, voice changes, genders and numbers. Therefore, known methods have their own advantages and disadvantages for application on documents written on the Serbian language. The lack of adequate Serbian-Serbian morphological dictionaries in digital form is one more drawback in solving problems for documents written on the Serbian language.

After that, a method for automatically generating a thesaurus for one corps documents in the Serbian language, as well as the results obtained by its application have presented. At the end, the conclusions and directions for further research are given.

## 2 RELATED RESEARCH

This section provides an overview of the methods used for the automatic extraction of similar words from a variety of sources: a large corpus of documents, web and monolingual dictionaries. The goal of these methods is to automatically find synonyms. Most of the methods finds words that are similar to each other with vague notation of semantic similarity.

The most of these approaches have a main assumption that similar words are used in the same context. These methods differ in the way of defining context (document, text, or less or more complex grammatical context). They also use different method for calculating the similarity function. Vector space model and cosine similarity measurement has been used to extract information from a specific corps by Chen and Lynch [5]. Crouch in [6] proposed a method for the automatic construction of the thesaurus. This method is based on term vector model and terms discrimination values [7], and this approach is particularly suited for words that do not occur frequently. Then a significant Grefenstettes SEKSTANT system [8] uses partial syntactic analysis, and solutions similar to this are described in [9] and [10]. Hierarchical clustering provides a rich thesaurus of similar words and has been discussed in [11]. Sometimes, a combination of several techniques can improve the list of synonyms. General solution to this problem is a mixed method [12]. Tourney dealt with Web as a corpus in [13], while the method for creating graphs that uses a thesaurus and dictionary has been described in [14].

The latest discoveries which use mathematical methods in creating thesaurus are useful for us, given the absence of morphological Serbian-Serbian language. One such method is described in the work [15]. It is particularly significant contribution of WordNet for Serbian language. The problems of synonyms in Serbian language for WordNet were engaged in the works of authors [16-18]. It should also be noted that in the experimental analysis has not started from scratch, but used part of the system for eGovernment for inserting documents described in [19].

## 3 METHOD DESCRIPTION

As noted earlier, unlike the English, the Serbian language has a lot of specifics rules like the voice changes and cases, verbal changes and numbers. All this makes the word to be in a different form from its base, although all this words form has the same semantic meaning. So, finding synonyms is essential in order

to effectively search documents written in Serbian language.

Serbian language is also lexically rich, and without morphological vocabulary available in electronic form, this similarity problem is even more difficult. There are online solutions that are incomplete, such as the project Vokabular [20], but the number of words contained here are so small that it is still practically unusable.

For automated extraction of similar words characteristic for a large body of documents, it is necessary to represent any document like set of key words (words and phrases that describe the content of the document). Automated extraction of keywords is a way of finding words that carry the meaning of the document. The form in which they appear in the document does not affect their semantic meaning. In order to find keywords in the document, the document first have to pass all levels of normalization.

The process of normalization of the document include: removal of stop words and no-letter signs, arranging numerical data, abbreviations, and splitting document on sentences, and then reducing all the words to their basic form (lemmatization in our case). Keywords have been extracted using the algorithm described in [19]. Most frequently occurring words have been extracted to the set of basic keywords. Extracting additional keywords has been done by calculating the deviation occurrence of specific words in the same sentence with the words in the basic set of keywords. The word that occurs more often with one specific keyword  $q$  seems to be the part of the key-phrase or new keyword. The significance of the deviation was assessed by chi-square test. The expected value has been calculated as the product of  $n_w$  (total number of words in sentences in which the word  $w$  appears), and  $p_q$  (the sum of the total number of words in sentences where the keyword  $q$  appears divided by the total number of words in the document) (Formula 1).

$$(1) \quad \chi^2 = \sum_{g \in G} \frac{(freq(w, q) - n_w * p_q)^2}{n_w * p_q}$$

Here is  $freq(w, q)$  the number of sentences in which the word  $w$  appears together with keyword  $q$  [21]. This method is significant because the extracting keywords have been done at the level of the document, without the need for multiple documents belonging to different corps. Once keywords, which are terms in a thesaurus, have found, those terms must be explained by stating the terms that are linked with them in a given corpus. These related terms, synonyms, have been found in the following manner: for each keyword in corpus (get keywords that appear in most documents in corpus) takes the vector of words that commonly appear next to it. Calculate the similarity between these vectors for all keywords using cosine similarity measure and Jaccard similarity coefficient. Thus, synonyms are declared as any keywords which vectors are the most similar. Pseudocode algorithm for creating thesaurus is shown below.

---

**Algorithm 1: Finding related terms for key terms in a thesaurus**

---

```
1: key1[] ← keywords in corpus;
2: n ← the number of key words taken;
3: i ← 0; j ← 0;
4: While i ≤ n do
5: while j ≤ n do
6: i_key_pair[] ← words that occur the most frequently with the key1[i];
7: j_key_pair[] ← words that occur the most frequently with the key1[j];
8: cos_coef[i,j] ← cosine similarity between i_key_pair[] and j_key_pair[];
9: jacc_coef[i,j] ← Jaccard coefficient for i_key_pair[] and j_key_pair[];
10: i++; j++;
11: end while
12: end while;
13: key2[] ← repeat steps 1-12 for the same documents with 4-gram cut
14: keywords[] ← key1[] intersection key2[]
```

---

Applying this algorithm on prepared but not normalized documents, results can be the words that are meaningful synonymous but also a derivation of the same word, so it may be possible to declare a synonym word as a case form of key terms.

## 4 EXPERIMENT RESULTS

The corpus of documents over which the experiment was carried out is a set of documents whose content consists of the Law on Higher Education. Figure 1 depict a part of the result of applying the intersections of two methods over the documents. First method presents results without document normalization and second method presents results with 4-gram normalization. By crossing the results, only those keywords that have emerged as candidates in both methods and only their associated words that have occurred by using both methods were taken, all this in order to maximize the accuracy of the result. Using the method described above, as the most important terms of the corpus, the following words were selected: "zakon", "rad", "program", "opšti" and "nastavnik". The table presents representative words of certain groups of related terms isolated using automated thesaurus. The first three words in the table are those that received more related words. Since this is a small number of documents, cosines coefficient values are high and equalized, but by adding new document to corpus, thesaurus is dynamically changing and obtaining more precise results.

By applying the algorithm, the last two words in the table produced by a small number of related words, but words are thereby significantly related to each other (eg. "opšti", "akt"). Hence we conclude that the words associated with fewer words with which occur together in sentences can determine the associated words more accurately. The first method did not use normalization, while the second method used normalization by discounting the first four letters

CSC - Cosine Similarity Coefficient JC - Jacquard Coefficient	Results without document normalization		Results with 4-gram normalization	
	CSC	JC	CSC	JC
<b>Related words</b>				
<b>Thesaurus entry</b>	<b>zakonom</b>		<b>zako</b>	
Skladu	0.8176	0.4577	0.791	0.4487
Obavlja	0.8281	0.3298	0.8603	0.4581
<b>Thesaurus entry</b>	<b>rada</b>		<b>ada</b>	
Skladu	0.7746	0.4388	0.7564	0.4403
Okviru	0.798	0.4677	0.8625	0.5313
Nastavnika	0.7702	0.487	0.7533	0.4846
Uslovi	0.7531	0.3497	0.8183	0.4329
<b>Thesaurus entry</b>	<b>program</b>		<b>prog</b>	
Pravo	0.7838	0.4095	0.8111	0.4453
Delatnost	0.7605	0.4948	0.7695	0.4054
<b>Thesaurus entry</b>	<b>opštim</b>		<b>opšt</b>	
Aktom	0.8106	0.3953	0.8718	0.5
<b>Thesaurus entry</b>	<b>nastavnika</b>		<b>nast</b>	
Rada	0.7702	0.3478	0.7533	0.3769

Figure 1: The result of applying the intersections of two methods over the documents

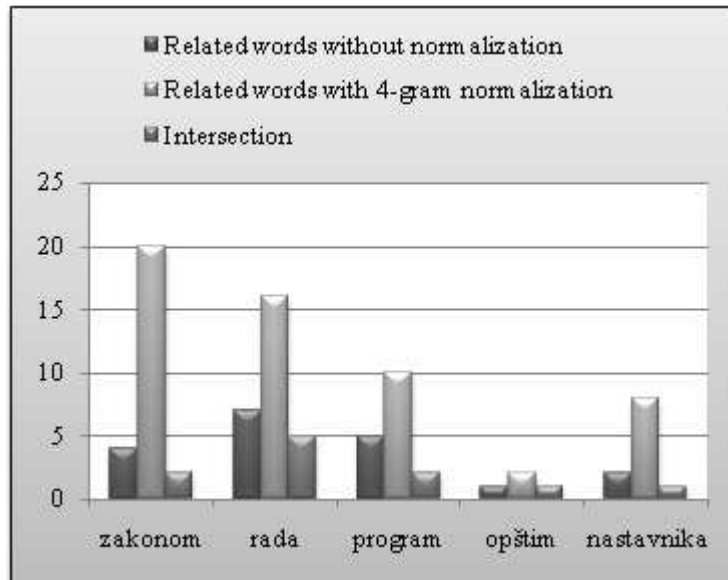


Figure 2: The number of related words in both methods and in their intersection

[22]. Thus, prepositions and one word derivatives incurred by adding a suffix were eliminated.

Figure 2 shows the amount of reduction of related words when you make a cross section of both applied methods. The results indicate that the normalization done by cutting off the first 4 letters produce a larger number of synonyms, so it can be used in cases where it is essential to expand a set of related words and reduce the precision of the results.

## 5 CONCLUSION

There are a number of different methods for the automated detection of similar words. Some techniques rely on different text corpora and mathematical similarity measures, while others use graphs and monolingual dictionaries.

Compared with hand-made dictionaries, computer-made lexicons have advantages such as their ease of construction and maintenance. Also, depending on the context, the term "similarity of the words" may or may not include the concept of synonyms, almost synonyms, antonyms, hyperonyms and so on.

The presented method use sentence as a context in which similar words exists. A function of the similarity between the words is calculated by comparing the word vectors with the chosen words appearing the most in a given context. The proposed method is based on mathematical methods and can be used without any normalization of documents and dictionaries. Normalization is improving the results, but the degree of automation of the process decreases. Subject to further work will be improving the algorithm using n-gram analysis and other forms of document normalization.

## ACKNOWLEDGMENT

This work was partially funded by the Ministry of Education and Science of the Republic of Serbia after the project III-44 007.

## References

- [1] Radovanovic M., Ferlež J., Mladenčić D., Grobelnik M. and Ivanović M., *Extending the IST-World database with Serbian research publications*, in Proceedings of IS2006, 9th International Multiconference on Information Society, pp. 251-254, Ljubljana, Slovenia, 2006.
- [2] Tan, P-N., Steinbach, M., Kumar, V., *Introduction to Data Mining*, Pearson Education, 2006.
- [3] Berry, M.W., Kogan, J., *ext mining: Applications and theory*, ISBN 978-0-470-74982-1, Wiley, 2010.

- [4] Berry, M.W. and Castellanos, ., Editors, *Survey of Text Mining: Clustering, Classification, and Retrieval*, Second Edition, September 30, Springer, 2007.
- [5] Chen, H. and Lynch, K.J., *Automatic construction of networks of concepts characterizing document databases*, IEEE Transactions on Systems, Man and Cybernetics Vol. 22, No. 1, pp. 885902, 1992.
- [6] Crouch, C.J., *An approach to the automatic construction of global thesauri*, Information Processing and Management, Vol. 26, No. 5, pp. 629640, 1990.
- [7] Salton, G., Yang, C.S. and Yu, C.T., *A theory of term importance in automatic text analysis*, Journal of the American Society for Information Science, Vol. 26, No. 1, pp. 3344, 1975.
- [8] Grefenstette, G., *Explorations in Automatic Thesaurus Discovery*, Kluwer Academic Press, Boston, MA, 1994.
- [9] Lin, D., *Automatic retrieval and clustering of similar words*, In Proceedings COLING, Montreal, Canada, August 1998.
- [10] Curran, J.R. and Moens, M., *Improvements in automatic thesaurus extraction*, In Proceeding ACL SIGLEX, Philadelphia, July 2002.
- [11] Pereira, F., Tishby, N. and Lee, L., *Distributional clustering of english words*, In Proceedings of ACL, Columbus, OH, June 1993.
- [12] Dietterich, T.G., *Ensemble methods in machine learning*, In Proceedings of MCS, Cagliari, Italy, June 2000.
- [13] Turney, P.D., *Mining the Web for Synonyms: PMI-IR versus LSA on TOEFL*, In Proceeding of ECML, Freiburg, Germany, September 2001.
- [14] Senellart, P., *Extraction of information in large graphs. Automatic search for synonyms*, Technical Report 90, Universite catholique de Louvain, Louvain-la-neuve, Belgium, 2001.
- [15] Nugumanova, A., Issabeava, D., Baiburin, Y., *Automatic Generation of Association Thesaurus Based on Domain-Specific Text Collection*, ISBN 978-80-87927-02-1, June 2014,
- [16] Krstev, C., Pavlović-Lažetić, G., Vitas, D., Obradović, I., *Using Textual and Lexical Resources in Developing Serbian Wordnet*, Romanian Journal of Information Science and Technology, Vol. 7, No 1-2, pp. 147-161, 2004.
- [17] Vujičić-Stanković, S., Krstev, C., Vitas, D., *Enriching Serbian WordNet and Electronic Dictionaries with Terms from the Culinary Domain*, Proceedings of 7th Global WordNet Conference, pp. 127-132, 2014.

- [18] Obradović, I., Krstev, C., Pavlović-Lažetić, G., Vitas, D., *Corpora Based Validation of WordNet Using Frequency Parameters*, Proceedings of the Second International WordNet Conference GWC-2004, Brno, Czech Republic, Masaryk University, pp. 181-186, 2004.
- [19] Crnišanin A., Avdić A., Marovac U., *Optimizacija dokumenata za njihovu bržu pretragu*, Nove informacione tehnologije za analitičko odlučivanje u biološkim, ekonomskim i sociološkim sistemima, State University of Novi Pazar, pp. 64-87, Proceedings, ISBN 978-86-86893-61-1, 2014.
- [20] *Srpski rečnik - Vokabular*, <http://www.vokabular.org/>, [Online] Januar, 2016.
- [21] Matsuo, Y., Ishizuka, M., *Keyword Extraction from a Single Document using Word Co-occurrence Statistical Information*, International Journal on Artificial Intelligence Tools, Vol. 13, No. 1, pp. 157-169, 2004.
- [22] Marovac U., Pljasković A., Crnišanin A., Kajan E. *N-gram analiza tekstualnih dokumenata na srpskom jeziku*, Proceedings of Telfor, pp. 1385-1388, 2012.