Available online at www.sciencedirect.com

ScienceDirect

Procedia Engineering 69 (2014) 296 – 303

Procedia Engineering

www.elsevier.com/locate/procedia

24th DAAAM International Symposium on Intelligent Manufacturing and Automation, 2013

From Patent Data to Business Intelligence – PSALM Case Studies

Zeljko Tekic[a]*, Miroslava Drazic[b], Dragan Kukolj[a], Milana Vitas[b]

[a]*University of Novi Sad, Faculty of Technical Sciences, Trg Dositeja Obradovica 6, Novi Sad, Serbia*
[b]*RT-RK Institute for Computer Based Systems, Narodnog fronta 23a, Novi Sad, Serbia*

Abstract

This paper describes PSALM, a recently developed software tool for business intelligence and its functionality through several case studies. Patent Search and Analysis for Landscaping and Management (PSALM) tool assembles patent data from publicly available data bases, collects and analyses bibliographic parameters of patents but also does text mining. High-dimensional data contained in the patent documents are transformed into much lower dimensionality space (2D or 3D), clustered and visualized. The PSALM functionality and usability is demonstrated through three case studies of analyzing, comparing and evaluating strengths and weaknesses of different patent portfolios.

Selection and peer-review under responsibility of DAAAM International Vienna

Keywords: patent data; PSALM; business intelligence; case studies

1. Introduction

Approximately 600 years ago first patents, in form of open letters with royal seal, were issued to glass-makers in Venice. Today, patent system promises to the owner the right to a temporary monopoly on a technical invention, in return for publication of that invention. Although it was not completely clear from the beginning, the patent system emerged as a tool for facilitating information dissemination and access to knowledge. For example, in return for a granted patent and a twenty years monopoly over the glass-making process previously unknown in England, John of Utynam (the recipient of the first known English patent in 1449), was required to teach his process to native Englishmen [1]. That same function of passing on information and new knowledge is still very important for the patent system.

* Corresponding author. Tel.: +381214852155; fax: +38121458133.
E-mail address: tekicz@uns.ac.rs

Selection and peer-review under responsibility of DAAAM International Vienna
doi:10.1016/j.proeng.2014.02.235

Rooted into patent's inherent characteristic – to disclose all details about protected products and processes, patents offer extremely valuable technical information. Some authors estimate that approximately 80% of all scientific and technical information can be found only in patent documents [2]. In addition to technical data, patent document provides legal as well as business and public policy relevant information. The availability of all these information inside patents offers a full spectrum of possibilities for using them in key areas of technology management including [3, 4]: competitors monitoring, technology assessment, the identification and assessment of potential sources for the external generation of technological knowledge and R&D portfolio management.

However, it is not easy to extract useful information from patents nor to track evidence about all patents that may be relevant. World Intellectual Property Indicators for 2012 [5] show that despite economic recession, around 2.14 million applications were filed and almost a million patents were issued around the world in 2011. With more than 65 million patent applications since the patent system was established, have been published; 7.88 million patents in force in 2011 and doubled number of granted patents over the last 15 years [5] it is possible to imagine how hard can be to track all interesting or potentially harmful patents. Other important barriers to the more efficient usage of patent information are: increasing number of pages per patent, difficult language used in patents and lack of ability to understand relations between patents.

Consequently, main stakeholders in R&D process – patent professionals, researchers and inventors, entrepreneurs, SMEs and commercial enterprises need help of software tools which will enable transformation of raw patent data into meaningful and useful information for business decision making. Various software tools have been developed in this field [2, 6]. They analyse individual patents as well as patent portfolios; retrieve patents and make basic statistics as well as visualize, map and landscape the same data. Most of these tools use statistical methods to analyze patent data in a specific period, and represent patent trends by various graphs and tables. In this paper we present PSALM [7, 8], recently developed software tool and demonstrate its functionality through several case studies.

The remainder of the paper is organized as follows. In Section 2 functional modules of PSALM and user interface are described, while in Section 3 PSALM functionality is demonstrated through three case studies. Finally, in Section 4 conclusion with a summary of our results and further research is outlined.

2. PSALM

All information found in a patent document is collected and verified according to internationally agreed standards. It is presented in a systematic manner, as a combination of structured and unstructured data. Technical information is derived from the description and drawings of the invention which disclose the technical details of the invention, illustrate working examples and show how to carry out the invention into practice. Legal information originates from the patent claims which define the scope of protection for the invention and from some of bibliographic data (priority date, date of filing, related patent documents, etc.). Finally, business and public policy-relevant information is derived from data identifying the inventor, date of filing, country of origin, etc.; and from an analysis of filing trends. The majority of information in patent document is given in the form of unstructured text. Only bibliographic data are structured. They are located on the front page and provide bibliographic information on the granted patent or patent application, which includes the document number, filing and publication dates, name of the inventors, assignees and addresses, etc.

PSALM (*Patent Search and Analysis for Landscaping and Management*) [7, 8] is a software tool designed to analyse both, structured and unstructured patent data. It consists of the following functional modules (Fig. 1): web robot, text clustering, multi-dimensional scaling, visualization, analysis of the IPC codes, extraction and display of citing and cited patents, progress report module, module for recording data in the CSV file, and evaluation of a patent. Modules are developed in programming languages Java and PHP, while database is developed in MySQL.

Software front-end (web robot) collects data on patents from publicly available data bases (USPTO and EPO), analyses their bibliographic parameters (like: title, inventor(s), applicant, date of application, priority date, country of publication, priority number, priority country, references cited by the patent, patents citing the patent, abstract, international patent classification) and translate unstructured data (free text in patent document) to structured form [7, 9]. The collected information is archived in the database for future use. The second module is text processing. Its main goal is to extract important attributes and keywords from a patent data structure. Text analysis includes analysis of patent text (abstract, description, claims or other data) using term frequency – inverse document frequency (tf-idf) as a weighting scheme for keyword extraction, although other methods can be used for classifying text streams by keywords [10]. The results have shown that analysis of claims offers the most accurate and relevant results [11]. Based on extracted keywords from the given dataset (collection of patent documents) the high dimensional matrix is formed. It is transformed into much lower dimensionality space (2D or 3D), maintaining the most similar structure to the original, using the multidimensional scaling (MDS) scheme. The output of the MDS is a 2-dimensional matrix which is used as an input for the third module – clustering. The reduced patent data space is clustered using unsupervised clustering technique in order to group the given unlabelled collection of patents into meaningful clusters. This approach enables to extract useful information from patents through the identification and exploration of keywords and key phrases of the textual data in the patents. There have been many different clustering approaches. Comparing the performances of four clustering techniques (i.e. k-means, the neural-gas, fuzzy c-means and ronn), it was shown that all have similar clustering performances and classification accuracy and thus any could be used in practical realizations of patent data analysis tools [12]. PSALM is based on fuzzy c-means clustering algorithm [12] where each patent has a degree of belonging to clusters, rather that belonging to just one cluster. Finally, the PSALM enables visualizations of high- as well as low-dimensional data. The high-dimensional data are visualized by mapping the documents and clusters in proportion to each other, i.e. creating patent maps. Documents with similar subjects appear close to each other in maps. This makes it very easy to locate the most developed areas in the technology. It also shows outliers in the data, patents that do not have much to the subject but are in the data by accident. Low-dimensional (structured) data are presented as bar charts and pie charts of bibliographic data and could also help in better understanding of the technology areas, changes in the technology development, company competiveness etc.

PSALM collects and stores patent data (access to the web page and download web page with the patent data; Parse the web page; Store data in database) within 2s (download/upload speed 26/1 Mb/s). TF-IDF processing time for group of 1800 patents is around 15 minutes, while MDS and visualization are done within 3s [7].

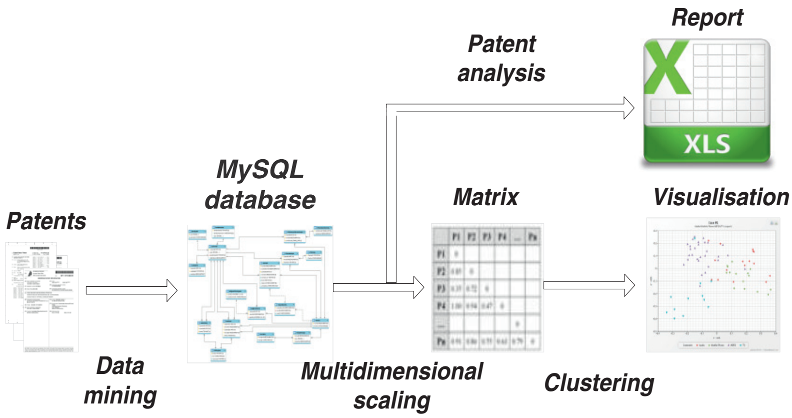

Fig. 1. Structure of the PSALM tool.

2.1. User interface

PSALM is a software tool developed to analyze a larger number of patents and to serve multiple networked users at the same time in server – client manner. The whole system is case-based, where each case is made of group of patents selected on basis of the users' defined criteria. Criteria for creating a new case can be based on: assignee, IPC codes and cited and citing patents. In addition to these criteria, the user can create unlimited number of criteria for selecting patents based on keywords and bibliographic attributes. Each case is unchangeable after creation. However, it is possible to create a new case with a different set of patents combining existing cases. Patents should be entered directly number-by-number (PID) or as list in .csv form.

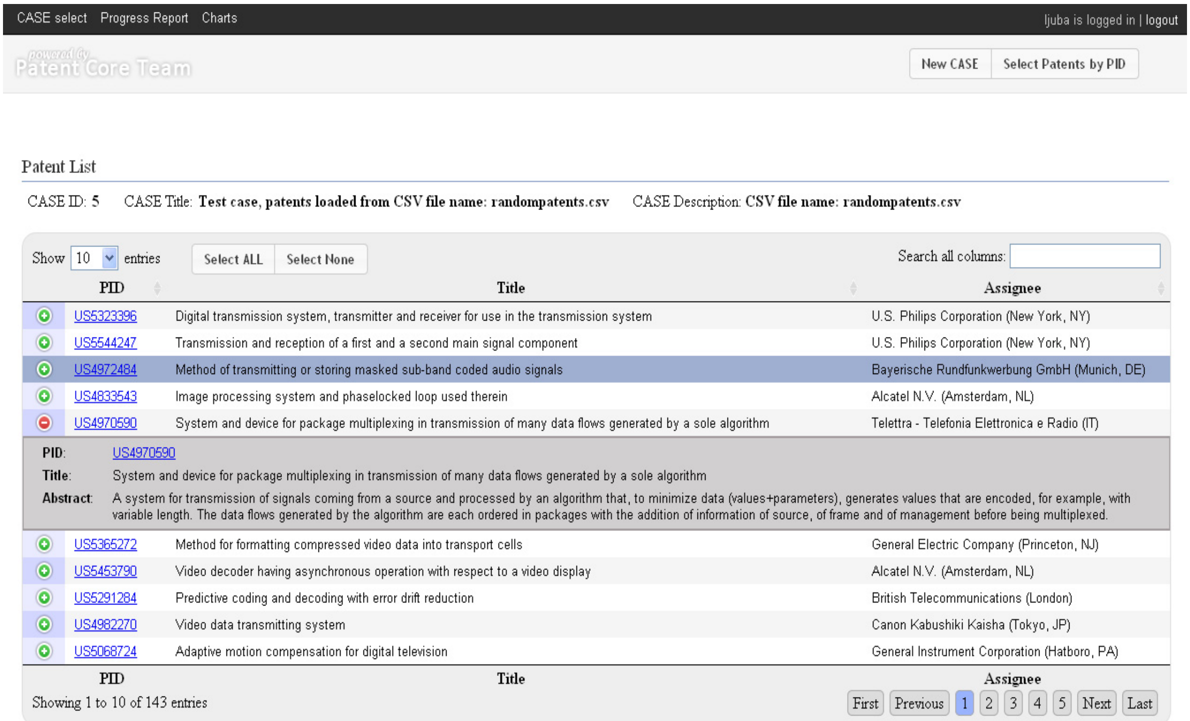

Fig. 2. PSALM user interface.

The user interface (Fig. 2) is built using PHP, HTML and JavaScript programming languages as well as JQuery JavaScript library, DataTables and HighCharts library for displaying the results of data processing.

3. Case studies

In this section the PSALM functionality is demonstrated. Analysis and evaluation of the company's patent portfolio strength are the tasks which re-occur in a daily work of a patent analyst. Therefore, such use cases are selected to illustrate the PSALM functionality.

3.1. Case #1

In the first case 147 US patents which belong to MPEG-2 essential patent portfolio were selected. A patent is essential to a standard, if making a product or using a method, complying with the standard, requires use of the patent. The task was to indicate strength of some companies in MPEG-2 field comparing essential patents and patents citing them. Fig. 3 shows specific areas in which two selected companies: LG (green triangles) and Toshiba (red squares) have technology advantages or disadvantages comparing with the set of essential patents (blue rhombi). From Fig. 3 it is possible to conclude that LG has strong position in audio coding and video transmission, while Toshiba is better positioned in coding/decoding digital signals. On the other hand, both companies are in good situation in areas of video coding/decoding and video compression. At the same time Fig. 3 verifies PSALM's ability to assemble patents into technology meaningful groups. Namely, these patents were first analysed by experts and clustered. Ellipses in Fig. 3 are placed additionally for the purpose of illustration only, to show satisfactory matching between the tool and human experts' results.

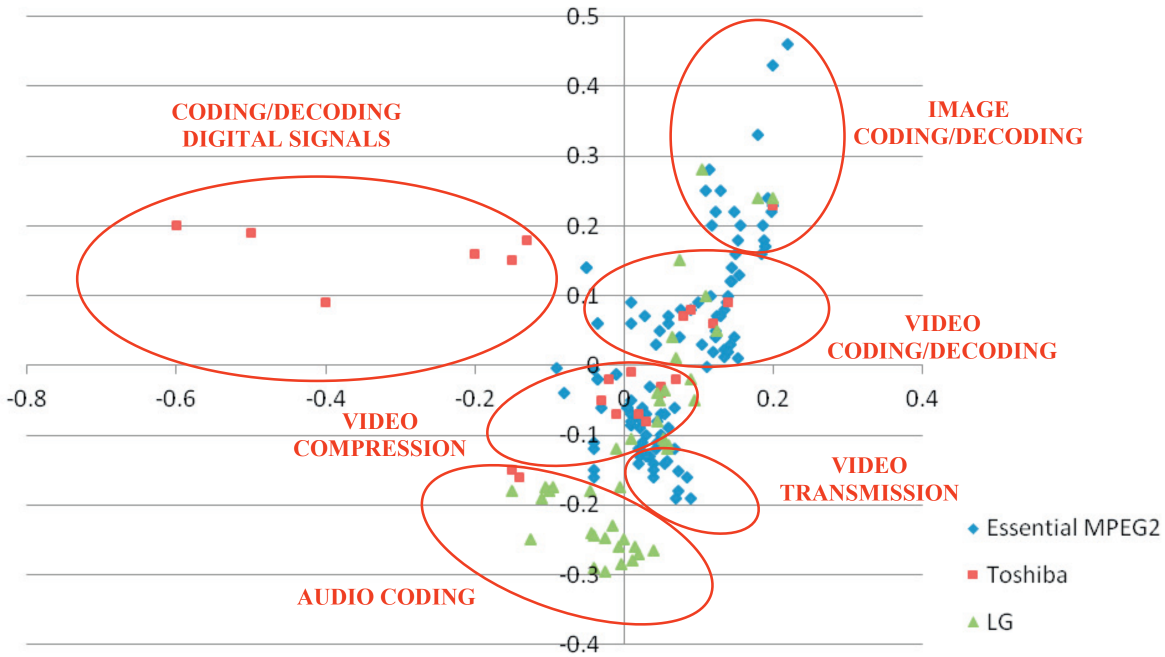

Fig. 3. Comparing MPEG-2 essential patents and companies' portfolios.

3.2. Case #2

The data set which was selected in the second case consists of 19 patents (further: original patents) which belongs mostly to technology field of distribution of multimedia content and represent the portfolio of one SME. The task was to find relevant companies and assess the strength of their portfolios in relation to portfolio of this SME.

Fig. 4. SME portfolio vs. Microsoft portfolio.

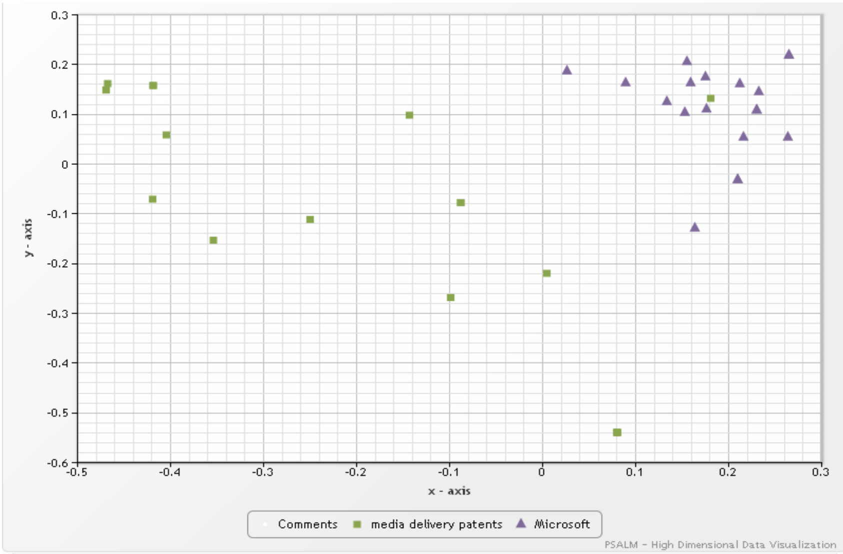

Fig. 5. SME portfolio vs. Microsoft portfolio, dominant IPC codes only.

Using the PSALM tool it was found that Microsoft has the highest number of patents among 115 patents which were citing original patents (forward citations) and which were cited by them (backward citations) indicating that it was the most active company in the field. Therefore, Microsoft was selected as a primary target for checking. Analyzing the original patents using clustering based on IPC codes, two most common IPC codes were detected (G06F21/00 and H04l9/00). Then all Microsoft patents containing both of these two codes were retrieved (19 patents in total) as well as all Microsoft patents containing at least one of these two codes (726 patents in total). Fig. 4 shows how 19 original patents match to 726 Microsoft patents, while Fig. 5 shows how 19 original patents match to 19 Microsoft patents.

It can be seen from the figures 4 and 5 that although the Microsoft has a large number of patents in the same technological area as the SME, these patents do not overlap in 2D space, which means that they are not closely related to each other. Namely, Microsoft patents are concentrated in one part of the 2D space, while the original 19 patents are located in the other part. Original patent which is the closest to the Microsoft patents in case two (the only green square among triangles at Fig. 5), is the closest original patent to Microsoft patents in case one as well (red diamond among densely spaced squares at Fig. 4). Additional (human) expertise proved that the nearest Microsoft patents are related to some encryption schemes for streamed multimedia content which is protected by rights management and not particularly related to enhancing copyright revenue, like the patents of SME. This was a way to verify the tool accuracy.

3.3. Case #3

In the third case, patents which are related to Android operational system are in focus. The task was to analyze patent litigations related to Android OS and from that perspective reflect on Google decision to buy Motorola Mobility. Searching through litigations related to Android OS between 2009 and 2012, 55 patents were detected [13]. Analyses done by the tool indicated that these 55 litigated patents cited 22 Motorola Mobility patents. Fig. 6 shows how 55 litigated patents match to 22 Motorola Mobility patents.

Analyses of detected and litigated patents revealed that Motorola's patents are relatively well distributed and related to patents which can harm Google. From that point, many who argued that Google decision to buy Motorola Mobility is partly rooted in its patent portfolio were right. On the other hand, Motorola does not have enough patents close to the patents under litigations, so it seems that Google will have to do several more purchases on the market to be in safer position.

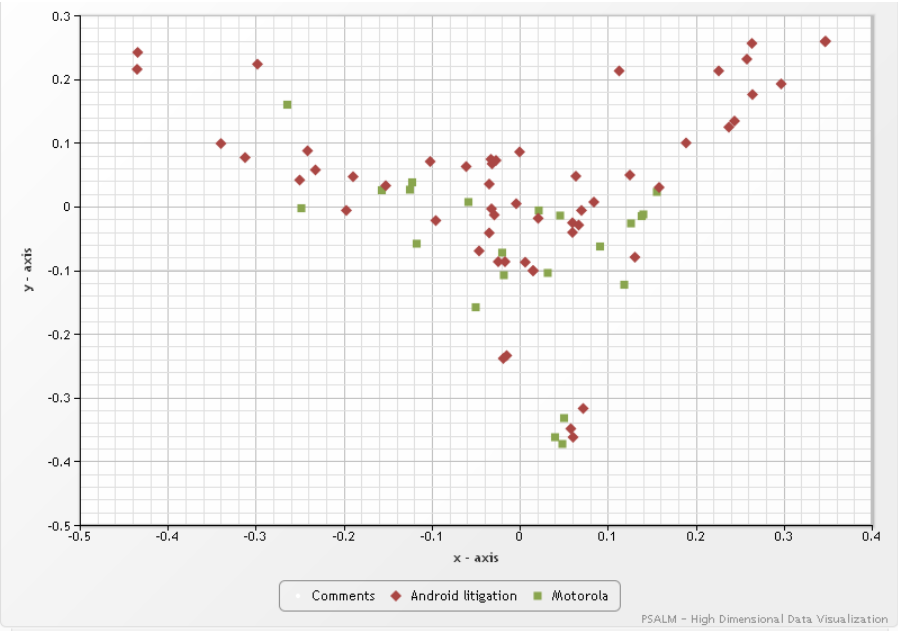

Fig. 6. Android (litigated) patents vs. Motorola Mobility patents.

4. Conclusion

In this paper we presented PSALM – a tool for patent data analysis and visualization developed by academics from University of Novi Sad and practitioners from RT-RK Computer Based Systems LLC. Its real power is in analyzing portfolios with a larger number of patents. This is demonstrated on three case studies of analyzing, comparing and evaluating strengths and weaknesses of companies' patent portfolios.

Patent data analyses will still be hard, time and manpower consuming experts' work, but PSALM could help professionals involved in IP management to focus their time and efforts on the most interesting and most promising patents, but also to save time in preliminary grouping them. For example, based on PSALM results it is easier to target technology weak areas or to select with higher probability patents interesting for infringement sues. Knowing which patents are interesting and why they are interesting is important especially for those who make decisions about usage and management of patents.

Results presented in this paper are results of current version of PSALM and further improvements are expected in the next period. The tool can be used to extract more meaningful data representation from the large set of patents. Further research will be directed towards tool improvement in text processing, using WordNET for comparing words in the text and SAO structures for text analysis. Also, future work will be concentrated on extending the test data set in order to further verify the results and improve data mining techniques, clustering and visualization modules.

Acknowledgements

This work was partially supported by the Ministry of Education, Science and Technology Development of the Republic Serbia under Grant number TR-32034, III-44009; and by the Provincial secretary of Science and Technology Development of Vojvodina Province under Grant number 114-451-2434/2011-03.

References

[1] Thomson Reuters, The History of Patents, available on: http://ip-science.thomsonreuters.com/support/patents/patinf/patentfaqs/history/; accessesed on 10 July 2013, 2013.
[2] L. Ruotsalainen, Data mining tools for technology and competitive intelligence, Espoo, VTT Tiedotteita – Research Notes 2451, 2008.
[3] H. Ernst, Patent information for strategic technology management, World Pat. Inf. 25:3 (2003) 233-242.
[4] A. Segev, J. Kantola, Identification of trends from patents using self-organizing maps. Expert Sys Appl. 39:18 (2012) 13235-13242.
[5] WIPO, World intellectual property indicators 2012 (WIPO Publication No. 941E/2012), WIPO, available at: www.wipo.int/export/sites/www/freepublications/en/intproperty/941/wipo_pub_941_ 2012.pdf/; accessesed on 15 July 2013, 2012 .
[6] H. Dou, V. Leveillé, S. Manullang, and J. M. Dou, Patent analysis for competitive technical intelligence and innovative thinking, Data Sci. J. 4:31 (2005) 209-237.
[7] Z. Tekic, D. Kukolj, LJ. Nikolic, M. Drazic, M. Pokric, M. Vitas, Z. Panjkov, D. Nemet, PSALM – Tool for business intelligence. *Proceedings of 35th MIPRO - International convention on information and communication technology, electronics and microelectronics,* Opatija, Croatian Society for Information and Communication Technology, Electronics and Microelectronics – MIPRO, 2012, pp. 1975-1980.
[8] Z. Tekic, D. Kukolj, LJ. Nikolic, M. Pokric, M. Drazic, M. Vitas, SMEs, patent data and new tool for business intelligence. Proceedings of 5th International Conference for Entrepreneurship, Innovation and Regional Development ICEIRD, Sofia, St. Kliment Ohridski University Press, 2012, pp. 855-863.
[9] LJ. Nikolic, D. Kukolj, M. Pokric, M. Drazic, M. Vuckovic, M. Vitas, Web robot – patent data acquisition software (in Serbian), Proceedings of 56th conference for electronics, telecommunications, computers, automation, and nuclear engineering – ETRAN, Etran Sociaty, Belgrade, 2012, RT 5.5, pp. 1-4.
[10] B. Yang, Y. Zhang, X. Li, Classifying text streams by keywords using classifier ensemble. Data Knowl Eng, 70:9 (2011) 775–793.
[11] M. Drazic, D. Kukolj, M. Vitas, M. Pokric, S. Manojlovic, Z. Tekic, Effectiveness of text processing in patent documents visualization, Proceedings of 11th International IEEE Symposium on Intelligent Systems and Informatics, SISY 2013, Subotica, 2013, pp. 287-291.
[12] D. Kukolj, Z. Tekic, LJ. Nikolic, Z. Panjkov, M. Pokric, M. Drazic, M. Vitas, D. Nemet, Comparison of algorithms for patent documents clusterization, Proceedings of 35th MIPRO - International convention on information and communication technology, electronics and microelectronics, Opatija, Croatian Society for Information and Communication Technology, Electronics and Microelectronics – MIPRO, 2012, pp. 1176-1178.
[13] M. Drazic, Contribution to the solution of automatic processing of patent documents, Master thesis, University of Novi Sad, 2012.