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.

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

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