Available online at www.sciencedirect.com

ScienceDirect

Procedia CIRP 56 (2016) 193 – 198

www.elsevier.com/locate/procedia

9th International Conference on Digital Enterprise Technology- DET2016 – "Intelligent Manufacturing in the Knowledge Economy Era

Extraction of Principle Knowledge from Process Patents for Manufacturing Process Innovation

Gangfeng Wang[a,b,*], Xitian Tian[b], Junhao Geng[b], Richard Evans[c], Shengchuang Che[a]

[a] Key Laboratory of Road Construction Technology and Equipment of MOE, School of Construction Machinery, Chang'an University, Xi'an 710064, China
[b] Institute of CAPP & Manufacturing Engineering Software, Northwestern Polytechnical University, Xi'an 710072, China
[c] Westminster Business School, University of Westminster, London, NW1 5LS, UK

* Corresponding author. Tel.: +86-029-82334586. *E-mail address:* wanggf@chd.edu.cn; wanggf@mail.nwpu.edu.cn

Abstract

Process patents contain substantial knowledge of the principles behind manufacturing process problems-solving; however, this knowledge is implicit in lengthy texts and cannot be directly reused in innovation design. To effectively support systematic manufacturing process innovation, this paper presents an approach to extracting principle innovation knowledge from process patents. The proposed approach consists of (1) classifying process patents by taking process method, manufacturing object and manufacturing feature as the references; (2) extracting generalized process contradiction parameters and the principles behind solving such process contradictions based on patent mining and technology abstraction of TRIZ (the theory of inventive problem solving); and (3) constructing a domain process contradiction matrix and mapping the relationship between the matrix and the corresponding process patents. Finally, a case study is presented to illustrate the applicability of the proposed approach.

Peer-review under responsibility of the scientific committee of the 5th CIRP Global Web Conference Research and Innovation for Future Production

Keywords: Process Innovation; Knowledge Management; Knowledge Acquisition; Patent Mining; Principle Knowledge

1. Introduction

Innovation has always been regarded as an important factor in maintaining the competitive advantage and market position of manufacturing companies, especially in the face of fierce competition often present in global markets. In recent years, theoretical methods and the application of innovative design have gradually become the common concern of academia and industry [1]. With the development of information and communication technologies, Knowledge Management (KM) and theoretical approaches to innovation [2], a new category of tools known as Computer-Aided Innovation (CAI) is being developed; these offer an effective way to assist designers to achieve creative inspiration and improve the efficiency of technological innovation. The goal of CAI is to support enterprises in effectively implementing a complete innovation process throughout the entire product life cycle; this includes fuzzy front end, product development, manufacturing, service and recycling, up to and including successful innovations in the marketplace [3]. Process innovation is a positive step in seeking to guarantee the delivery of product innovation and is also fundamental to the sustainable development of manufacturing [4-6]. As a branch of CAI, Computer-Aided Process Innovation (CAPI) can stimulate the creative thinking of process designers and help them to implement process innovation through the adoption of structured or systematic approaches [7].

Process innovation design is a structured innovative implementation process based on knowledge, and consequently formalized process innovation knowledge acquisition is crucial for CAPI. Process patents have become an important knowledge resource for process innovation design due to their innovative and practical features, but the inherent principle knowledge contained within patent text does not lend itself easily to the application of such knowledge in process innovation [8]. On the other hand, the contradiction

Peer-review under responsibility of the scientific committee of the 5th CIRP Global Web Conference Research and Innovation for Future Production
doi:10.1016/j.procir.2016.10.053

matrix of classical TRIZ (Theory of Inventive Problem Solving), formed through the analysis and extraction of several million technology patents, is used to point out the inventive principles that can be applied to solve technical contradiction of specific problems [9]. However the matrix is mainly suitable for product innovation due to the lack of specific process parameters and corresponding principles for innovation design.

In this research, we explore principle knowledge extraction from process patents for CAPI based on patent mining and knowledge management. An extraction framework of principle knowledge is described firstly. Then, we propose the methods of automatic classification of process patents and principle knowledge extraction based on patent mining. Finally, a case study of principle knowledge extraction from micro-cutting patents is illustrated.

2. Extraction framework of process innovation knowledge from process patents

From the systems thinking perspective, a specific problem-solving of process innovation mainly includes analysis and formulation of process problem, process conflict extraction and resolution, detailed design of process innovation scheme, evaluation and optimization of the scheme. Process innovation knowledge, which exists in the entire lifecycle of process innovation, is used to support process innovation activities correctly implemented and to produce new process knowledge. According to the knowledge demand and application of innovative design process, we divide process innovation knowledge into the following types: Problem Description Template, Process Contradiction Matrix (PCM), Manufacturing Scientific Effect, Innovative Scheme Instance, and Manufacturing Capability Description, etc [7]. Among them, PCM can provide the solution direction and innovative principle for technical conflict resolution of process problem-solving. And the process patents also contain innovative solutions and principles, so we can use the patent knowledge to build the PCM for CAPI. Here, we firstly establish a formal representation model of PCM, and then illustrate the PCM construction process based on patent mining.

2.1. Formal representation of process contradiction matrix

A technical contradiction arises when an attempt to improve certain attributes of a technical system leads to the deterioration of other attributes of that system [9]. Referring the classical TRIZ theory, we define the process contradiction as the phenomena of technical contradiction occurring in manufacturing systems. When the process contradiction hampers the realization of a process innovation goal, a process problem arises. In this paper, the parameters with contrary behavior characteristics are referred to as process contradiction parameters. The parameter which is expected to get enhanced or improved is called strengthening parameter, while the parameter which is expected to get reduced or downgraded is called weakening parameter. The combination composed of any one strengthening parameter and any one weakening parameter is called a process contradiction pair. Generally, each process contradiction pair will have several corresponding basic solving directions, namely the solving principles for process contradiction. Solving principles are the general laws to resolve those process contradictions. Here, a process contradiction matrix is used to represent the relationships between process contradiction pairs and the corresponding solving principles, as shown in Table 1. And it is defined as:

$$PCM = \bigcup_{\substack{i,j=1 \\ j\neq i}}^{n} PCM_{ij} = \bigcup_{\substack{i,j=1 \\ j\neq i}}^{n} \left\langle \left(\overrightarrow{Par_i}, \overleftarrow{Par_j} \right), Sp_{ij} \right\rangle \quad (1)$$

where $PCM_{ij} = \left\langle \left(\overrightarrow{Par_i}, \overleftarrow{Par_j} \right), Sp_{ij} \right\rangle$ stands for process contradiction unit, and n is the total number of process contradiction parameters. $\overrightarrow{Par_i}$ and $\overleftarrow{Par_j}$ represent strengthening parameters and weakening parameters, respectively. $Sp_{ij} = \left\{ Sp_{ij}^1, Sp_{ij}^2, \cdots, Sp_{ij}^k \right\}$ is a set of solving principles for a process contradiction, and k is the number of solving principles.

Table 1. The form of process contradiction matrix.

Weakening parameters → / Strengthening parameters ↓	Par_1	Par_2	Par_3	…
Par_1		Sp_{12}	Sp_{13}	
Par_2	Sp_{21}		Sp_{23}	
Par_3		Sp_{32}		…
…		…		…

2.2. Construction process of process contradiction matrix

The process contradiction matrix construction based on patent mining is a knowledge conversion process that maps the unstructured patent text into the structural innovation knowledge by using Natural Language Processing (NLP) technology [10]. As can be seen in Fig. 1, process contradiction matrix construction should be based on the classified process patents in a specific way firstly, and then process contradiction parameters and contradiction solving principles can be extracted respectively from the patents under the support of knowledge base, finally principle knowledge will be associated by backtracking the mining process. Thus construction process mainly consists of the following parts: process patents classification, process contradiction parameters mining, contradiction solving principles mining and principle knowledge association for process contradiction matrix.

Process patent documents need to be pre-processed before the data mining. We store the required parts of the patents, and form a process patent database having unified data format. Patent text generally has a relatively uniform format, for example US patents mainly have Title, Abstract, Claims, Background of the Invention, Summary of the Invention, Description of the Invention, etc. It needs to deal with a lot of contents for automatic classification of process patents and core innovation knowledge extraction from them with computer-aided technology. If we analyze the whole patent content, it will lead to a large amount of computation and prone to excessive interference information. At the same time, because of the particularity of the patent text, there is a phenomenon of repeated narration between the parts of the patent. Therefore, it is necessary to select the appropriate representative components from the process patent text. Through manual analysis of process patents, we found that the Title or Abstract of a patent generally include some feature words of process method, manufacturing object or manufacturing feature which can be used to distinguish the process fields. Moreover, by examining the different parts of the patents, we found that the description of process contradiction parameters most likely in Abstract, while Abstract and Summary of the Invention basically can reveal the adopted technical method or inventive principle for process problems. In this research, we choose Title, Abstract, Background of the Invention, and Summary of the Invention as the main information source of patents.

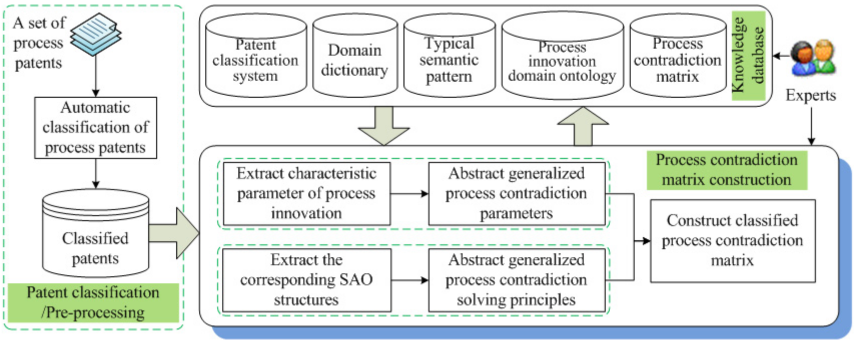

Fig. 1. Construction process of process contradiction matrix based on patent mining.

Semantic analysis of process patent mining is a process of natural language understanding, which requires several knowledge database supports and domain experts' participation. Among them, patent classification system stores the classification criteria. Domain dictionary constituting the concept hierarchy based on the hyponymy relation of domain concepts contains predefined characteristic parameters and their identification, as well as domain term abbreviations. It can be used to support concept standardization and attribute identification in the extraction of process contradiction feature parameters and Subject-Action-Object (SAO) structures. Typical semantic pattern database which stores the typical semantic pattern for identifying feature parameters can be used to match patent texts containing these parameters. Process innovation domain ontology describes the semantic information of concepts through the relationship between concepts. Based on the concept hierarchical structure, the self-defined ontology relations are added to this database, such as Associative-Relation, Part-of, and Cause-Effect. Process innovation domain ontology can be used to support the domain generalization of feature parameters and SAO structures.

3. Automatic classification of process patents

Because a process patent has its own specific innovation intention, in order to establish a process contradiction matrix with the process domain characteristics and universality, we need to classify process patents reasonably, and establish the set of classified patents. Process patents generally contain innovative technological solutions, i.e., creative application of specific process method to achieve the processing of specific manufacturing object and its manufacturing feature. Here, a classification criterion for process patents is established by taking process method, manufacturing object and manufacturing feature as the references, as shown in Fig. 2 (a). Besides, a process classification system and the corresponding classification system base are constructed according to the specific technological characteristics of manufacturing enterprises. For example, an aero-engine manufacturing enterprise could divide their manufacturing objects into blade, turbine and diffuser, process methods into milling, grinding and EDM, manufacturing features into cylindrical surface, hole and root plane.

Automatic classification of process patents can be regarded as two parts, classification learning process and innovation oriented process. And the learning process is divided into the training process and testing process, as illustrated in Fig. 2 (b). In the training process, a classifier is constructed according to the learning model of training patents; while in the testing process, testing patents are classified by using classifier and testing results will be fed back to the classifier in order to improve classification performance. The learning process is a process that requires constant feedback and improvement. Thus the specific automatic classification of process patents mainly includes the following steps: 1) Preparing sample database of process patents, 2) Choice of representative components for patent classification, 3) Feature extraction and selection of process patents, 4) Establishment of feature representation model, 5) Construction and test of classifier.

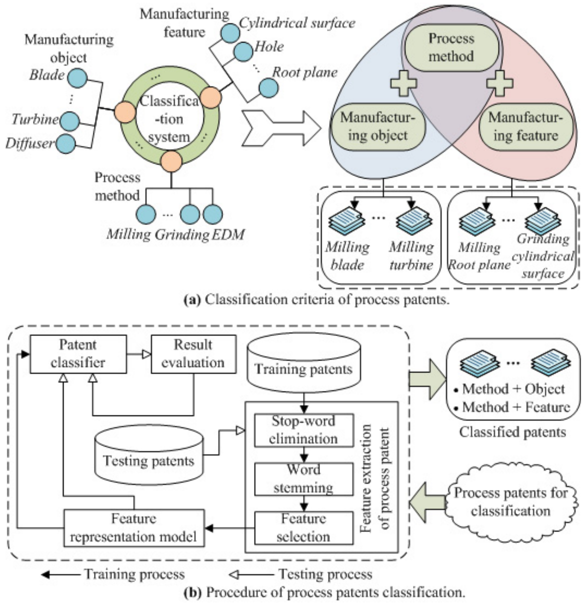

Fig. 2. Process patents classification based on specific criteria.

4. Principle knowledge extraction based on patent mining

As revealed in Fig. 1, we know that process contradiction matrix can be established by extracting the principle knowledge, which consists two parts of knowledge: process contradiction parameters and the corresponding contradiction solving principles. In this section, the method of extraction and association for these knowledge is explored based on patent mining.

4.1. Mining method of process contradiction parameters

The basic mining procedure of process contradiction parameters is shown in Fig. 3. Firstly, the characteristic parameters are extracted from classified process patents under the support of typical semantic pattern database, then these characteristic parameters will be clustered according to their attributes, subsequently generalized process contradiction parameters can be obtained based on process innovation domain ontology database.

In this research, a typical semantic pattern based characteristic parameters extraction algorithm is used to successively extract characteristic parameters from classified process patents and add the attribute identification into these parameters. Process patent text can be separated by statements and added the Part-Of-Speech tagging by using POS Tagger [11]. Thus each sentence of patents can be matched with typical semantic patterns, and then core terms or phrases of the successful matching sentences will be extracted as the candidate characteristic parameters.

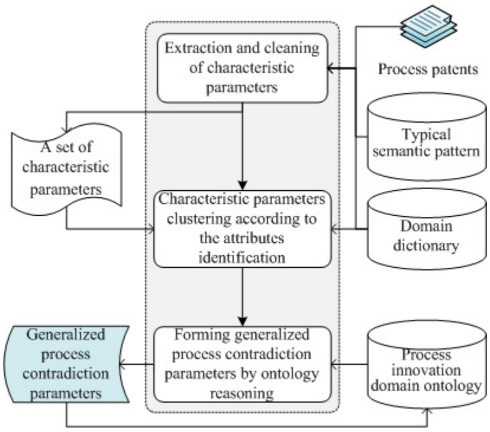

Fig. 3. Mining procedure of process contradiction parameters.

In order to improve the efficiency and accuracy of mining process, it is necessary to filter out irrelevant and repetitive parameters in the same patent text. Here, referencing the concept hierarchy definition of WordNet, we construct a domain dictionary for process innovation to realize word normalization and semantic disambiguation. For example, ‘rotary ultrasonic drilling’ and ‘RUD’ both represent a type of ultrasonic drilling. In the domain dictionary, we define the two words as hyponym of ‘dictionary-ultrasonic drilling-noun’.

These candidate characteristic parameters can be clustered according to the attribute identification of parameter properties which express from two aspects: system feature and solving feature. Because parameter properties have many dimensions, the characteristic parameter may belong to more than one cluster region. The clustering results can be roughly divided into two types: 1) belonging to a property identification, i.e., system feature or solving feature; 2) simultaneously belonging to properties of system feature and solving feature. For example, characteristic parameters, ‘feed rate’, ‘cutting speed’, ‘cutting depth’ and ‘cutting force of rake face friction zone’ are clustered together as they all belong to machining process parameters; while ‘dimensional accuracy’, ‘shape accuracy’ and ‘position accuracy’ are clustered as they are all the measured accuracy parameters.

Process innovation domain ontology stores the concepts and relations of domain knowledge model, thus generalization result can be obtained according to the core words of characteristic parameters. The generalization process of process contradiction parameters also needs domain experts’ participation due to the high abstraction. In each set of the clustered characteristic parameters, these parameters having the common hypernym will be generalized by searching the domain ontology relationships. And the semantics of upper concept is more abstract and shared than the underlying concepts, and the underlying concept is more specific and more close to the specific application. For example, clustered characteristic parameters, ‘mechanical efficiency’, ‘machining efficiency’, ‘processing time’ and ‘clear corner efficiency’ are generalized as ‘production efficiency’; parameters, ‘fixture’, 'measuring instruments', 'tool' and 'auxiliary tool' are generalized as 'technical equipment'.

4.2. Mining method of process contradiction solving principles

From the point of view of TRIZ, SAO structure is essentially a function model and basic semantic unit that represents the solution of problem-solving [9, 12]. Thus key information for solving process problems and resolving process contradictions can be found in the SAO structures extracted from process patent text. The basic mining procedure of process contradiction solving principles is shown in Fig. 4. Firstly, the specific process solutions of classified process patents are expressed as extracted SAO structures by using SAO sentence analyzer, then these SAO structures will be clustered according to their attributes, subsequently the clustered SAO structures can be generalized under the support of process innovation domain ontology database and experts participation.

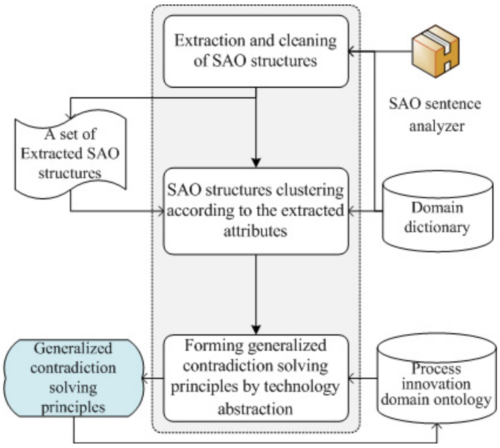

Fig. 4. Mining procedure of process contradiction solving principles.

Since the related studies and tools of NLP have been conducted to extract SAO structures from sentences, such as Stanford parser, Knowledgist and PAT-Analyzer, we extract SAO structures based on these methods and tools. For example, a SAO structure 'manipulator-convey-profiled rod', is extracted from the abstract of US7707705 using Stanford parser. However the expression these SAO structures are not standardized, we also conduct data cleaning on them, which includes low frequency clipping, concept standardization of Subject and Object, and word sense disambiguation. After the cleaning is completed, each process patent can be expressed as the corresponding set of SAO structures.

Because one SAO structure only expresses a specific process solution, rather than a generic technical function or solution, these SAO structures need to be clustered and analyzed based on the extracted attributes. According to the affiliation features, relevant SAO structures containing the common generalized process contradiction parameter are clustered together. According to the contribution features, the SAO structures having the common solution objective are clustered. For example, patents US8560113 and US7933679 can be clustered due to they have the same process contradiction pair 'material removal efficiency ↔ tool life'; patents US8560113 and US20090148296 can be clustered considering their innovation objective 'decreasing processing times'.

Similarly, clustered SAO structures also need to be abstracted by technology abstraction which is a basic approach of TRIZ [9]. Here, we carry out technology abstraction by using an abstract SAO model of process ontology, which consists of the noun concept and the fact type (represent a relationship between 'Subject' and 'Object'). The fact type is composed of partative and effect fact types. Generally, partative facts describe the inclusion relationship between two noun concepts, and effect facts describe how the subject concept affects the object concept. Partative SAO structures express the relationships between products or technologies, while effect SAO structures describe the realization of technical functions, i.e., the relationships of 'problem-solution' in which the action-object (AO) states the problem and the subject (S) forms the solution. For example, 'Magnet heater-raises-temperature of fluid' expresses a complete solution to the problem, while 'raises-temperature of fluid' shows the technical functions that need to be realized, and 'Magnet heater' is the solution. Therefore, partative SAO structures are helpful to analyze the composition of process system, and effect SAO structures can further form the generalized contradiction solving principles. Thus, with the experts' participation, generalized contradiction solving principles can be formed for a kind of process contradictions resolution.

4.3. Principle knowledge association for process innovation

During the process of knowledge extraction, we embed the identification for process contradiction parameters and the corresponding solving principles from certain process patents. Then a process contradiction matrix can be established according to the mapping relationship between process patents and process contradiction knowledge. A process contradiction matrix describes a formal representation of principle knowledge by process contradiction parameters and the corresponding contradiction solving principles, and achieves the data association between principle innovation knowledge and process patents. Thus, the combination of principle knowledge and knowledge source will help designers to stimulate their creative thinking in innovation.

5. Case study

A case study of principle knowledge extraction of micro-cutting from process patents is performed to investigate the effectiveness of the proposed method.

Firstly, we manually analyze more than 500 patents and select 237 copies as the sample set of patent classification based on patent classification standard system of an aviation manufacturing enterprise. The sample set is divided into training set and testing set according to the ratio of about 2:1. In stage of principle knowledge extraction, some process patents related turbine processing are downloaded from United States Patent and Trademark Office (USPTO), and Support Vector Machines (SVM) is selected as classifier, thus 97 turbine milling patents are obtained by classification. Subsequently, 2208 original characteristic parameters are extracted from classified process patents by using the typical semantic pattern based characteristic parameters extraction algorithm. After data cleaning, 534 parameters were obtained. Meanwhile, 2935 SAO structures are extracted by using Stanford parser, and 928 structures are left after data cleaning.

Through the extraction and analysis of characteristic parameters and SAO structures, the generalized process contradiction parameters and corresponding solving principles are formed, and then process contradiction matrix units are established successively under the support of experts' participation. By analyzing the mining results and patent data source, we find that patent mining process is more time saving and convenient than manual analysis. However, the recall rate of mining result is still relatively low, so the knowledge bases need to be enriched, especially the typical semantic pattern. In addition, based on the computer-aided patent mining and manual analysis, we initially formed a micro-cutting process contradiction matrix (as shown in Fig. 5) which has been used in process problem solving of a micro-turbine by combining with other innovation knowledge of CAPI.

Weakening parameters / Strengthening parameters	1.Work-piece structure	2.Process efficiency	3.Tooling complexity	4.Fixture clamping force	5.Fixturability of workpiece	6.Manufactu-ring quality	…
1.Workpiece structure	+	6，11	1，6，7，9，12	1，8，10		9	
2.Process efficiency	3，6	+	2，8，9	1	6，8，11	1，12	
3.Tooling complexity	4，5，9	3，8，12	+	11，12	3		
4.Fixture clamping force		1，7	5，10，12	+	4，7	3	
5.Fixturability of workpiece	2，5			7，12	+	1，6，13	
6.Manufactu-ring quality	12	3，7	9	5		+	
……							

Fig. 5. A part of micro-cutting process contradiction matrix.

6. Conclusion

By indicating the important role of innovative principles for innovation design, especially in relation to CAPI, an extraction framework and approach of principle knowledge from process patents has been proposed. A formal representation model of process contradiction matrix has been built which can reasonably organize principle innovation knowledge. On this basis, several key technologies, automatic classification of process patents, mining method of process contradiction parameters and corresponding solving principles are studied. The research describes a mapping framework of unstructured patent text to structured process innovation knowledge, which can provide support for process innovation design through the use of patent knowledge. In the future, we will focus on the understanding of how to improve the automaticity and efficiency of knowledge mining.

Acknowledgements

The authors are grateful to partial support of the National Natural Science Foundation of China (51305042), Specialized Research Fund for the Doctoral Program of Higher Education (20130205120004), Industrial Research Project of Science and Technology Department of Shaanxi Province (2016GY-003), Natural Science Foundation of Shaanxi Province, China (2016JQ5107), and Fundamental Research Funds for the Central Universities (310825161005, 310825151039). We also would like to express special thanks to Y.-G. Liu and X.-B. Wang of NWPU for their time and involvement in system development, data collection and testing.

References

[1] Cavallucci D, Leon N. Computer-supported innovation pipelines: Current research and trends. Computers in Industry 2011; 62(4):375-6.

[2] Xu J, Houssin R, Caillaud E, Gardoni M. Fostering continuous innovation in design with an integrated knowledge management approach. Computers in Industry 2011; 62(4):423-36.

[3] Leon N, Cho SK. Computer aided innovation. Computers in Industry 2009; 60(8):537-8.

[4] Fan Z, Chen L, Zhang Z. On the principle and case studies of the technology innovation. China Mechanical Engineering 2002; 13(4):350-3.

[5] Duflou JR, D'hondt J. Applying TRIZ for systematic manufacturing process innovation: the single point incremental forming case. Procedia Engineering 2011; 9:528-37.

[6] Geng J, Tian X. Knowledge-Based Computer Aided Process Innovation Method. Advanced Materials Research 2010; 97-101:3299-302.

[7] Wang G, Tian X, Geng J, Guo B. A knowledge accumulation approach based on bilayer social wiki network for computer-aided process innovation. International Journal of Production Research 2015; 53(8):2365-82.

[8] Cascini G, Russo D. Computer-aided analysis of patents and search for TRIZ contradictions. International Journal of Product Development 2007; 4(1):52-67.

[9] Ilevbare IM, Probert D, Phaal R. A review of TRIZ, and its benefits and challenges in practice. Technovation 2013; 33(2-3):30-7.

[10] Li Z, Tate D, Lane C, Adams C. A framework for automatic TRIZ level of invention estimation of patents using natural language processing, knowledge-transfer and patent citation metrics. Computer-Aided Design 2012; 44(10):987-1010.

[11] Santorini B. Part-of-speech tagging guidelines for the Penn Treebank Project (3rd revision), 1990.

[12] Choi S, Park H, Kang D, Lee JY, Kim K. An SAO-based text mining approach to building a technology tree for technology planning. Expert Systems with Applications 2012; 39(13):11443-55.