International Journal of Innovative Computing, Information and Control
Volume **14**, Number **3**, June **2018**

ICIC International ©2018 ISSN 1349-4198
pp. 1133–1139

RESEARCH ON CONSTRUCTION OF PATENT DYNAMIC TECHNOLOGY EFFECT MATRIX

Yao Liu[1,2] and Yi Huang[1]

[1]Institute of Scientific and Technical Information of China
No. 15, Fuxing Road, Beijing 100038, P. R. China
liuy@istic.ac.cn

[2]Fujian Provincial Key Laboratory of Information Processing and Intelligent Control
Minjiang University
No. 1, Wenxian Road, Shangjie University Town, Fuzhou 350121, P. R. China

Received August 2017; revised December 2017

Abstract. *The analysis of patent effect matrix can reveal the relationship between technology and effect in the patent literature, resolve obscure information content and potential technical characteristics, and find out technical focus and blank spots to avoid technology minefields. The construction of effect matrix is the prerequisite for the analysis of patent effect matrix. Currently, it needs the efforts of the intelligence analyst, the domain expert or the enterprise technician to interpret each patent document in detail so as to understand the technical and utility information. In view of this, this paper puts forward a method of constructing dynamic technology effect matrix based on domain ontology, information extraction, and semantic annotation and so on. And the implementation and application in the chemical field show our method can achieve good results.*
Keywords: Effect matrix, Patent analysis, Ontology, Semantic annotation

1. **Introduction.** Patent technology effect analysis is regarded as a method of patent analysis by most companies and researchers. The effect matrix can intuitively reflect the development of technical areas, top assignees, countries, and regions information. Through the analysis, we can get the appropriate technology research areas and development opportunities as well as avoiding technical traps. In general, effect matrix construction can be divided into the following steps: technical and functional classification construction, patent literature interpretation analysis, patent literature summary analysis, data summarization, and effect matrix generation. At present, most studies are focusing on patent effect analysis, and there are few studies in patent effect matrix construction [1,2]. Jun et al. built effect matrix by clustering to serve technical forecast, but the method of obtaining technical words and functional words was not introduced [3]. Cheng used International Patent Classification (IPC) to obtain technical words, and constructed the patent effect matrix [4]. Todirascu et al. built a knowledge system, assigned quantitative importance with the help of domain experts, identified and extracted related technical and functional words based on grammar and domain knowledge [5]. Wang et al. proposed a method to mine the patent literature with lexical database and subject indexing, but the extraction of technical and functional words was not mentioned [6]. Huang and Hsu extracted technical and functional information from patent claims to identify important technologies in cloud computing [7]. He et al. proposed a method to use semantic role labeling to extract effect terms form sentences that express the patent advantage from the DII (Derwent Innovations Index), and select high-frequency words from IPC as technology words to create patent matrix [8]. Methods of term extraction

in patent literature include: statistical-based, rules-based, domain knowledge (including IPC, CPC, Japan File Index, F-term, etc.) based, patent grammar structure-based methods, but the classification criteria are not unified, and the existing classification cannot be used directly to build the matrix, for they are either too broad, or only applicable to specific areas [9-11]. It is hard to adjust to the diversity of patent literature, and to meet the in-depth patent annotation needs. For instance, IPC emphasizes holistic classification, and it is difficult to reflect the specific technical point; F-term does not distinguish between technology and effect terms, many subject terms can be placed in a number of areas, and it needs further manual interpretation and refinement. In addition, for a very specific technology and effect theme, the structure derived from the patent classification cannot completely cover these topics.

In view of the issues above, this paper puts forward a method of automatic construction of patent dynamic effect matrix. Different from other patent effect matrix analysis methods, such as using MPEST (Material, Personality, Energy, Structure, Time) technology perspective, TEMPOS (Treatment, Effect, Material, Process, Product, Structure) map, IPC classification, Japan File Index, Japan F-term classification and other technical categories to obtain related terms, our method is to extract words from each and every patent as technical and functional term candidates, and mine and determine the appropriate application characteristics, statistical characteristics of the patent literature based on domain ontology and semantic annotation technology in an automatic, controllable and adaptive way to form a dynamic and interrelated patent knowledge base, for the generation of patent technology effect matrix.

The rest of this paper is organized as follows. Section 2 describes the idea and framework. Section 3 describes the key technologies. Implementation and application are shown in Section 4. Finally, conclusion and future work are given in Section 5.

2. **Idea and Framework.** To be more specific, we use a domain ontology containing technical and functional classification to facilitate term extraction and semantic annotation with natural language processing tools; at the same time, we use patent literature to renew and enrich the concepts, properties, and relations of the ontology. In this way, the components, technologies, and utilization of patent literature as well as the matrix generation pattern are established by using the ontology knowledge. Based on the interrelated knowledge network, with a click on a term on the technology axis or the effect axis of the matrix, a new matrix is automatically generated to provide the related information a researcher or company concerns, such as technology gaps analysis and technical research and development trend analysis. The overall framework is shown in Figure 1.

3. **Key Technologies.** With the design idea and framework, this paper conducts research in technical and functional system construction, semantic annotation, and dynamic technology effect matrix generation.

3.1. **Technical and functional system construction.** With the classification of various technical fields defined by experts, there is a certain degree of domain knowledge generalization, but the degree of technical words refinement is not enough. It often misses some important technical words, especially the new technical words, which weakens the value of the effect matrix analysis. In order to solve the problem of technical terms discovery and refinement, based on our previous research [12,13], the method is adopted: fully utilizing the existing structured data and semi-structured data to build domain ontology, together with patent semantic framework (including patented verbs, grammatical features, semantic roles, etc.), the semantic relations between the terms are well revealed and extracted to define the technical and functional terms of each patent; when some

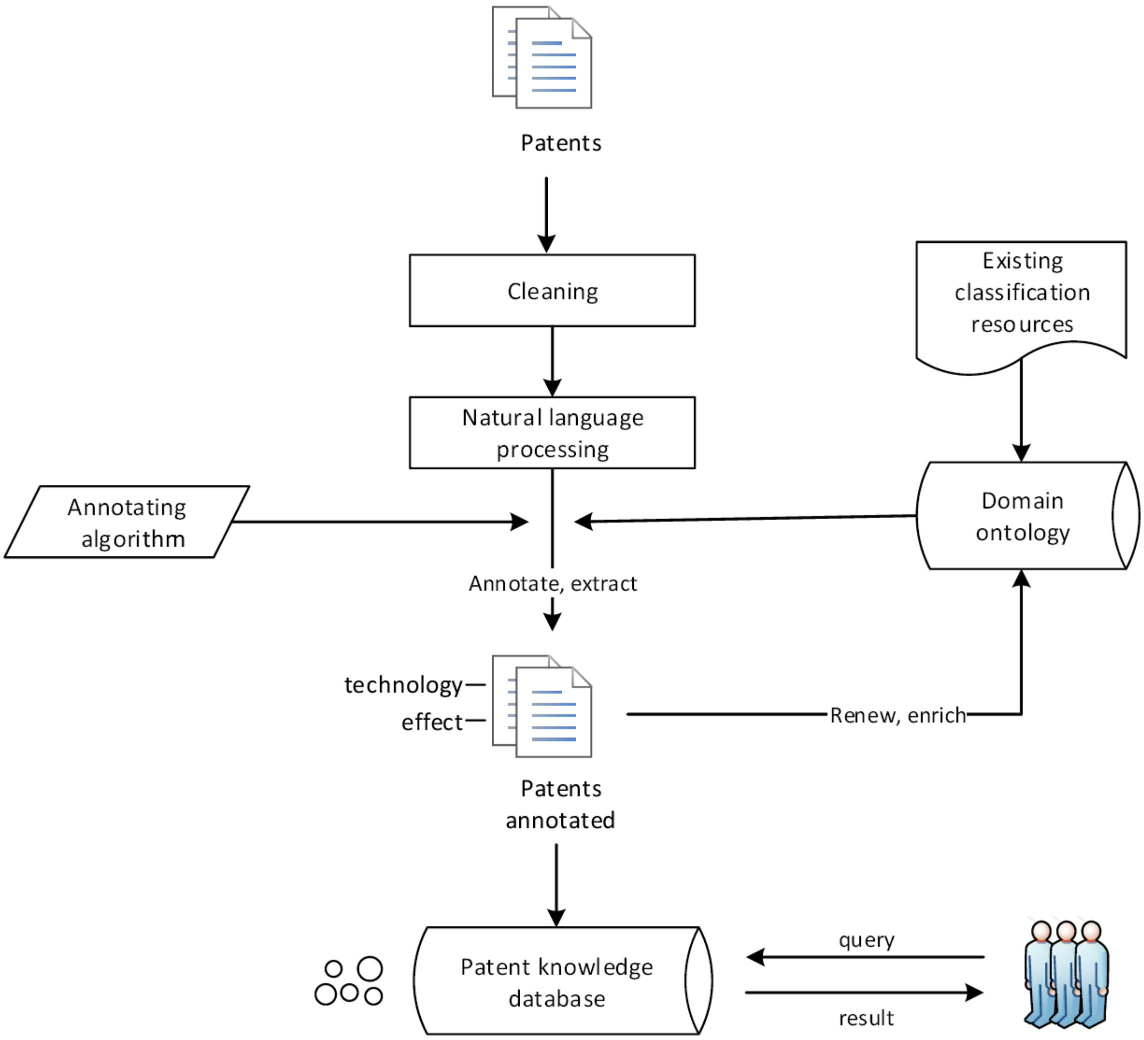

Figure 1. Design idea and framework

new words are not in the framework, they can be automatically added to the hierarchical structure and established connection with patent literature; at the same time, all these data can be used to renew and enrich the domain ontology.

3.2. **Semantic annotation.** In order to fully understand the content of patent information, based on our previous studies [14,15], we propose to use natural language processing technology to conduct in-depth annotation of the patent text. Semantic annotation is the process of organizing semanticization of document resources: a set of semantic concepts is extracted according to the frequency, position, and relationship of concepts in the patent text. The resource used in semantic annotation is domain ontology. Domain ontology provides definition, property, and relation between concepts. Patent text is being cut into different sizes of text fragments. The basic idea of semantic annotation is that patent is not indexed according to the whole content: on the one hand, the whole patent is too broad to describe the concept, which may involve many aspects of the concept; on the other hand, the semantic information of the whole patent may involve multiple concepts, not easy to generalize to the basic semantic content. The semantic annotation is carried out in two dimensions to further understand the meaning of a patent and extract terms: the paragraph level and the sentence level. The former is used for indexing the whole content, and the latter is used for better extraction of relevant concepts to better understanding the meaning of patent text and accurate term extraction for the patent text.

3.3. **Dynamic technology effect matrix generation.** With the classification structure being established, the corresponding model of patent component, technology and utility being established, the semantic annotation process being conducted, a dynamic, interconnected patent knowledge base is constructed. The related technical and functional terms can be easily retrieved and a basic structure of effect matrix is formed. With the statistics of patent documents on each and every intersection of the matrix, the patent effect matrix is generated.

4. **Implementation and Application.** We used Windows and MyEclipse as development environment, JAVA as programming language, MongoDB as database, to complete the system development. Chemical patents are chosen to verify our proposed method. We take material, structure, and process as technology term, performance improvement, easy operation, energy saving, time-saving, etc. as effect term. It is in complete accord with the cognition upon patent technology effect matrix by mainstream research [16,17]. A chemical ontology with technology term classification and effect classification, containing 19,222 concepts, 13,544,251 semantic relations is being automatically constructed, and the chemical patent literature is being annotated. We choose 10 chemical material terms and 10 common effect terms from the chemical ontology and conduct the queries in the patent database. The rate of the sum of the common effect terms with the total is chosen as evaluation score. Firstly, we execute the queries without considering the structure relation between the terms, and the result is shown in Table 1.

Table 1. Patent database query result 1

Technology (material) / Effect	food additives	potassium sorbate	sorbic acid	sodium citrate	citric acid	inorganic acid	sulfuric acid	nitric acid	hydrochloric acid	phosphoric acid
safety	**1065**	875	1405	1211	6982	**620**	18198	7013	8557	16899
clean	**120**	201	367	597	2988	**469**	7810	2863	2595	5534
fresh-keeping	**316**	519	808	205	1818	**23**	1435	517	425	1248
purify	**114**	102	163	313	1851	**292**	10035	3997	3223	4501
stable	**982**	1352	2612	3836	20843	**2413**	54226	24420	23876	54047
effective	**1206**	1452	2621	2563	16354	**2209**	49214	17471	22166	46110
environment-friendly	**281**	362	634	1346	6782	**712**	26970	9389	8856	19839
low cost	**625**	680	1133	1505	8526	**1274**	32802	15721	15311	22859
good effect	**143**	290	548	603	3072	**195**	9699	2913	3034	7368
easy to use	**963**	1105	1702	1223	7690	**521**	17705	6993	7333	14455
...	**...**	...	...	...	...	**...**	...	...	...	...
total	**7924**	7623	12470	14960	96665	**14889**	304971	130292	133250	267573
rate	**73.38%**	91.01%	96.17%	89.59%	79.56%	**58.62%**	74.79%	70.07%	71.58%	72.08%

As can be seen from Table 1, sorbic acid gets the highest rate, and the average rate of all the technology terms is 77.68%. It can be concluded that the 10 common effect terms basically cover the effect of the technology terms, while these terms have a rather close relationship with the top rank effect of sorbic acid compared with other technology terms. However, food additives, the hypernym of potassium sorbate, sorbic acid, sodium citrate, citric acid, and inorganic acid, the hypernym of sulfuric acid, nitric acid, hydrochloric acid, phosphoric acid, get a comparatively low rate. These relations are defined in the chemical ontology, which can be easily retrieved and calculated with proper configuration. Then, we conduct all the queries again, and the result is shown in Table 2.

Table 2. Patent database query result 2

Technology (material) / Effect	food additives	potassium sorbate	sorbic acid	sodium citrate	citric acid	inorganic acid	sulfuric acid	nitric acid	hydrochloric acid	phosphoric acid
safety	**11538**	875	1405	1211	6982	**51287**	18198	7013	8557	16899
clean	**4273**	201	367	597	2988	**19271**	7810	2863	2595	5534
fresh-keeping	**3666**	519	808	205	1818	**3648**	1435	517	425	1248
purify	**2543**	102	163	313	1851	**22048**	10035	3997	3223	4501
stable	**29625**	1352	2612	3836	20843	**158982**	54226	24420	23876	54047
effective	**24196**	1452	2621	2563	16354	**137170**	49214	17471	22166	46110
environment-friendly	**9405**	362	634	1346	6782	**65766**	26970	9389	8856	19839
low cost	**12469**	680	1133	1505	8526	**87967**	32802	15721	15311	22859
good effect	**4656**	290	548	603	3072	**23209**	9699	2913	3034	7368
easy to use	**12683**	1105	1702	1223	7690	**47007**	17705	6993	7333	14455
...	**115054**	...	...	...	...	**616355**	...	...	...	...
total	**139642**	7623	12470	14960	96665	**850975**	304971	130292	133250	267573
rate	**82.39%**	91.01%	96.17%	89.59%	79.56%	**72.43%**	74.79%	70.07%	71.58%	72.08%

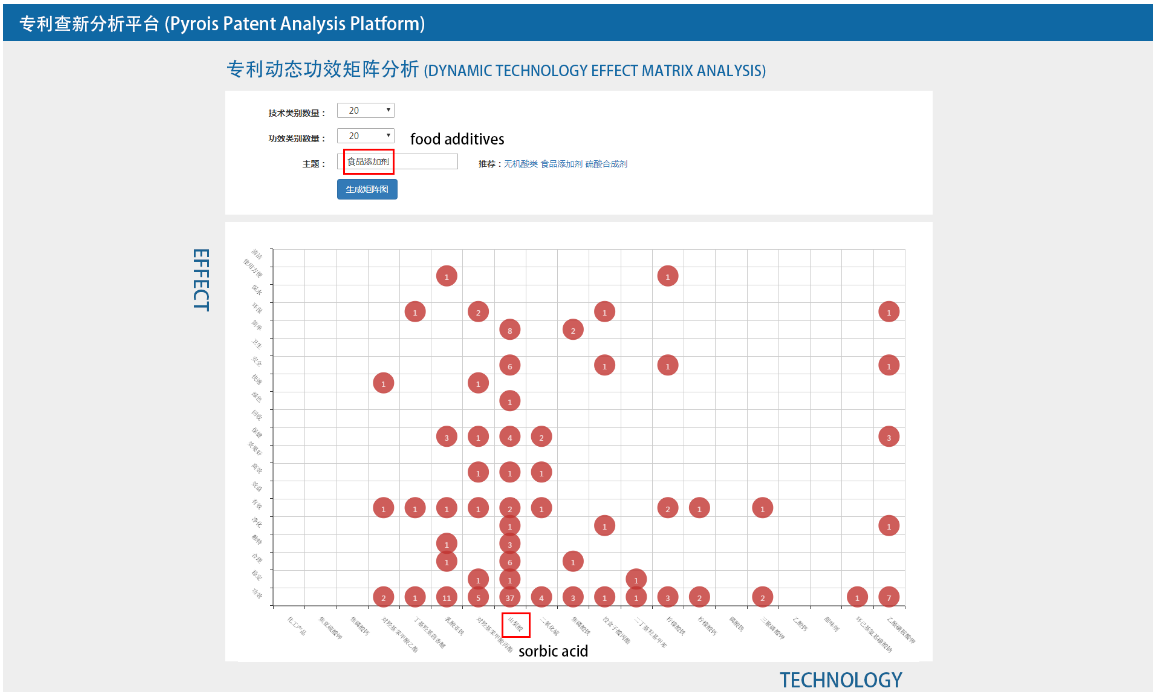

Figure 2. Patent effect matrix analysis result 1

As can be seen from Table 2, the rate of hypernym words (i.e., food additives, and inorganic acid) gets a significant increase in both number and rate. By taking the relations in domain ontology into account, we can get a more reasonable result. It can be also concluded that our method is with flexible capability, and the result is highly related to the concept structure, property, and relation presented in the domain ontology.

Some application is also conducted. After choosing the number of terms on technology axis and effect axis, entering the query word, for example, "food additives", a patent effect matrix is automatically generated.

As can be seen from Figure 2, the x-axis is the technology axis, the y-axis is the effect axis, and circle with a number is relative statistic and distribution information of related patents. We can click any term in either axis, for example, "sorbic acid", and a new

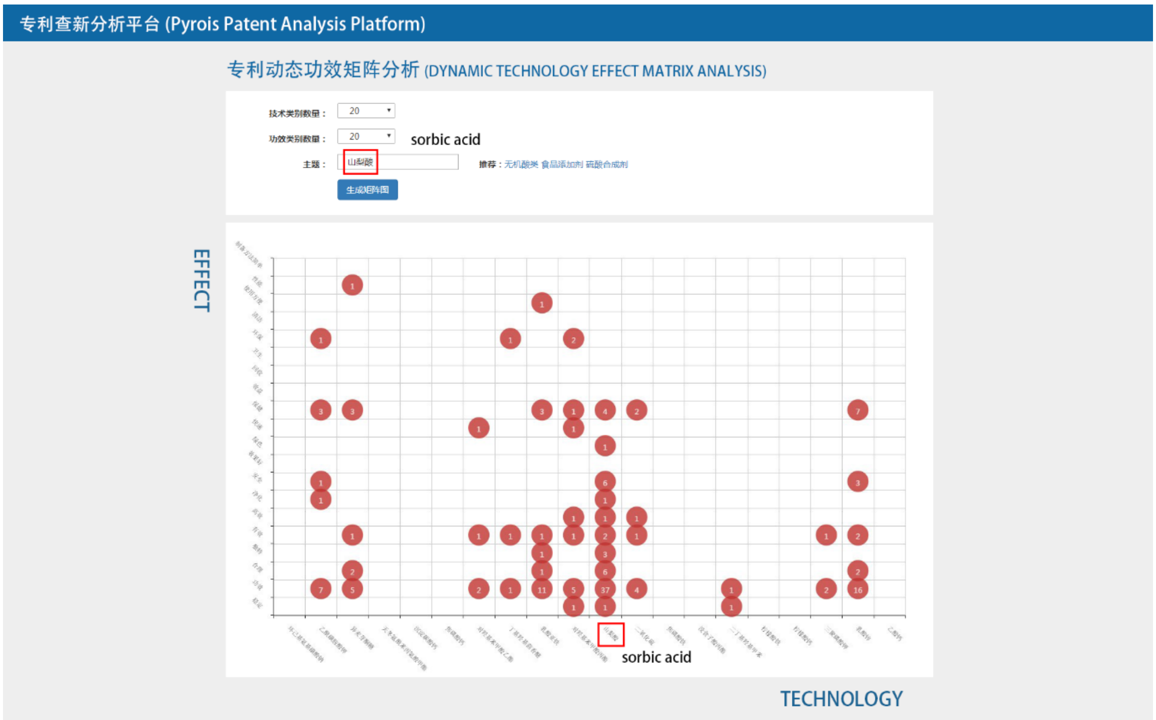

FIGURE 3. Patent effect matrix analysis result 2

technology effect matrix is being generated shown in Figure 3. The terms on x-axis are replaced with those terms closely related to sorbic acid, and the terms on the y-axis are replaced with those top effect terms of the technology terms accordingly.

5. Conclusions. In order to effectively address the problem of patent effect matrix generation under the era of big data, this paper puts forward a method of constructing dynamic technology effect matrix based on domain ontology, information extraction, and semantic annotation. Different from other patent effect matrix analyses, we extract terms from each and every patent as technical and functional term candidates, and with domain ontology and semantic annotation technology to mine and determine techniques, functions, application characteristics, statistical characteristics of the patent literature for the formation of patent dynamic effect matrix while renewing and updating information of the domain ontology. And the implementation and application in the chemical field show our method can achieve good results. In the future, we intend to proceed along two lines in parallel: on the one hand, to integrate other innovative resources such as journal articles and proceedings into the system; on the other hand, to broaden its application fields other than the chemical field.

Acknowledgment. This work is supported by National Key Project of Scientific and Technical Supporting Program No. 2013BAH21B02 and Open Fund Project of Fujian Provincial Key Laboratory of Information Processing and Intelligent Control (Minjiang University) No. MJUKF201739. The authors also gratefully acknowledge the helpful comments and suggestions of the reviewers, which have improved the presentation.

REFERENCES

[1] J. Yoon and K. Kim, Detecting signals of new technological opportunities using semantic patent analysis and outlier detection, *Scientometrics*, vol.90, no.2, pp.445-461, 2012.

[2] A. Abbas, L. Zhang and S. U. Khan, A literature review on the state-of-the-art in patent analysis, *World Patent Information*, vol.37, pp.3-13, 2014.

[3] S. Jun, S. S. Park and D. S. Jang, Technology forecasting using matrix map and patent clustering, *Industrial Management & Data Systems*, vol.112, no.5, pp.786-807, 2012.

[4] T.-Y. Cheng, A new method of creating technology/function matrix for systematic innovation without expert, *Journal of Technology Management & Innovation*, vol.7, no.1, pp.118-127, 2012.

[5] A. Todirascu, L. Romary and D. Bekhouche, Combining syntax and ontologies for information extraction, *Terminology and Knowledge Engineering – TKE'02*, 2002.

[6] L. Wang, D. Zhang, X. Zhang, X. Yang and M. Wu, Using the subject automatic indexing to generate technology effect matrix, *Modern Library and Information Technology*, no.5, pp.80-86, 2013.

[7] J. Y. Huang and H. T. Hsu, Technology-function matrix based network analysis of cloud computing, *Scientometrics*, vol.113, no.3, pp.1-28, 2017.

[8] Y. He, Y. Li and L. Meng, A new method of creating patent technology-effect matrix based on semantic role labeling, *International Conference on Identification, Information, and Knowledge in the Internet of Things*, pp.58-61, 2015.

[9] S. Choi et al., A fact-oriented ontological approach to SAO-based function modeling of patents for implementing function-based technology database, *Expert Systems with Applications*, vol.39, no.10, pp.9129-9140, 2012.

[10] A. Shinmori, M. Okumura, Y. Marukawa et al., Structure analysis of Japanese patent claims using cue phrases, *Journal of Geophysical Research Biogeosciences*, vol.119, no.6, pp.1216-1230, 2014.

[11] J. Yoon and K. Kim, TrendPerceptor: A property-function based technology intelligence system for identifying technology trends from patents, *Expert Systems with Applications*, vol.39, no.3, pp.2927-2938, 2012.

[12] Y. Liu, X. Chen, Z. Sui, Y. Hu and Q. Zhao, Research on semantic method of library resources' organizing, *ICIC Express Letters*, vol.5, no.4(A), pp.1011-1017, 2011.

[13] Y. Liu, H. Shi, D. Zheng and Y. Huang, Study on semantic annotation for professional literature, *ICIC Express Letters, Part B: Applications*, vol.5, no.5, pp.1383-1389, 2014.

[14] Y. Liu, Z. Zhang and Y. Huang, Research and development of semantic annotation platform for scientific literature, *ICIC Express Letters*, vol.10, no.7, pp.1787-1794, 2016.

[15] Y. Liu, Y. Li and Y. Huang, Research on semantic and syntactic analysis of patent literature, *ICIC Express Letters*, vol.10, no.2, pp.471-477, 2016.

[16] H. Nonaka et al., Extraction of the effect and the technology terms from a patent document, *International Conference on Computers and Industrial Engineering*, pp.1-6, 2010.

[17] S. Tejswi, S. K. Jain and R. S. Lokhande, Synthesis of alkyl aromatic compounds, AACs: Forecasting emerging technology through patent analysis, *International Conference on Management of Intellectual Property Rights and Strategy*, 2014.