



# **iJRASET**

International Journal For Research in  
Applied Science and Engineering Technology



---

# **INTERNATIONAL JOURNAL FOR RESEARCH**

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

---

**Volume: 7      Issue: III      Month of publication: March 2019**

**DOI: <http://doi.org/10.22214/ijraset.2019.3363>**

**[www.ijraset.com](http://www.ijraset.com)**

**Call:  08813907089**

**E-mail ID: [ijraset@gmail.com](mailto:ijraset@gmail.com)**

# Patent Analysis

Tejaswini O. Deshmukh<sup>1</sup>, Sakshi S. Parwani<sup>2</sup>, Shubhangi M. Tiple<sup>3</sup>, Parag G. Satav<sup>4</sup>, Mr. Prof. S. S. Sagane<sup>5</sup>

<sup>1, 2, 3, 4, 5</sup>Dept. of Computer Science & Engineering, P. R. Pote (Patil) Education & Welfare Trust's Group of Institutions College of Engineering & Management, Amravati, Maharashtra, India

**Abstract:** *The future of R&D, it is important to have an innovative idea. The main subject lies on which technology stream to choose which has long run benefits.*

*The above subject requires understanding of current technology and research trends.*

*The proposed project focuses on the learning patent topics and analyzes them for generating the co-occurrence network which defines the current technology streams for patents and suggest the innovative patent ideas.*

**Keywords:** *Patent Search, Patent Analysis, Text Preprocessing, Visualization, Scientific Literature, Naive Bayes Approach, KNN.*

## I. INTRODUCTION

In today's world Patents are the critical asset for any innovative company, and business which can increase global competition, aligning in the business strategies along with the IP strategies.

Patent informatics defines the science of searching, analyzing and presenting patent information to identify relationships and trends

Over the past few decades patent searching has changed dramatically from paper based access to instant online access, from library catalogue systems to internet search systems, from partially indexing to fielded and full text indexing to multi-modal indexing

According to Bonino et al [1], patent analysis can be divided into following main categories:

Search Within this section there are a number of separate types of search assignments which require two types of information like patents and non-patent to be retrieved and analysis in order to achieve a particular goal/work task.

The Analysis of patents is usually performed by an agent to notify users about new incoming patent information to keep users abreast of the latest developments.

The main focus the requirements associated with the search tasks like,

- 1) *State of the Art (SOA)*: identify patents for the purposes of a general review. Sometimes they are mention as Landscaping or/ and Technology survey.
- 2) *Novelty (NOV)*: classification of patents and non-patents may influence the patentability as an idea or an invention.
- 3) *Patentability (PAT)*: Here it gives the patent application and also ensure its novelty.
- 4) *Infringement (INF)*: recognize patents or its applications which cover the product or activity and are also force.
- 5) *Due Diligence (DD)*: analyze strengths, weaknesses and scope of IP rights.

Analysis The analysis of patents can be broken into two main types:

- a) Micro analysis of individual patents.
- b) Macro analysis of a group of patents.

The analysis is normally performed to evaluate the Intellectual Property (IP), to map and chart the IP, to introduce trends and also to recognized new areas of potential to exploit.

Literature review presents multiple different research works that have been carried out for good results in patent retrieval and analysis task or by using standard approaches at some different stages of the patent analysis process.

## II. LITERATURE SURVEY

Lupu et al. [3] give an in-depth account of current techniques for patent retrieval. Whereas it also contains the approach like Visualization to retrieval, it does not cover the entire range of visualization approaches to patent analysis.

Börner [5] focuses on mapping approaches, i.e. specializations of scientific communities often based on citation data. She provides a comprehensive overview including hand-drawn examples.

Our focus, in contrast, is broader, comprising all types of visualization methods that provide benefits for the analysis of scientific documents and patents.

*A. Regression Analysis*

Chan et al. mention about that Generalised Autoregressive Conditional Heteroscedasticity (GARCH) model and Glosten, Jagannathan and Runkle's (GJR) model can be used in time-varying volatility of patent ratio. Simultaneously, they use four countries as research objects, are Canada, France, Germany and Japan. The experiment results provide some evidence to support the impact that between these four countries to exist the presence of interdependent in the conditional variance of the patent growth rate [12].

Alok Khode et al [6] considering the various techniques and frameworks available and their limitations, there is a lot of scope in the field of patent retrieval techniques which makes room for further research to be taken up in this domain.

Longhui Zhang ET AL [7] introduces the recent development in the field of patent mining, it is still far away from being explored in research area. To assist patent analysts and attentive readers gain an over view of patent mining, it thus provide a detailed summary of present research done along this direction. This survey, outline the different technical trend in patent mining.

*B. Patent and Stock Market*

Yu et al. find that in the study of stock price movements' analysis, the numbers of patent of enterprise have better significant explanatory capability than R&D expenditure, because every patent data is unique. Patent except to as a key factors on research and development (R&D) expenditure in enterprise performance evaluation, it also can provide diversity information about enterprise innovation activities. The experiment results show that using the numbers of patent can add value in stock price movements' literature analysis [11]. Lin mention about that companies maintain the advantage of technologies' trend through innovation activities and creation activities, these two factors are be quantified by forward citation indicator and Import new technologies' indicators respectively, its stock price will present higher growth rate. The results show that more successful innovation activities will let companies' stock price have higher growth [8].

*C. Patent Analysis and Patent Quality Indicators*

Hirschey et al. through a series experiment to explore that the relationship between the value correlation of patent activity and the scientific value of patent information, and the results show that using scientific value indicators based on citation related information can as patent quality indicators, and they find that when they through patent quality related information to conduct communication successfully, the relationship between R&D expenditure and stock right value in more successful companies Tend to more consistent [9].

Kapoor et al. also point out some researchers have proposed a lot of value and quality indicators for patent to Identify and find out its potential value and quality, among them, citation related indicators are the most common evaluation indicators [10]

Table: - Various patent analysis techniques and approaches

Techniques:	Query formulation. Query expansion. Summarization. Relevance feedback.
Models:	Vector Space Model (VSM). Semantic Based Processing. Latent Semantic Analysis(LSA). Language Model. Weighting Techniques. Probabilistic Model etc.
Others:	Bibliometric methodology. Data Mining. Text Mining. Database Management Tools like OLAP. Citation Analysis.

### III. PROPOSED SYSTEM

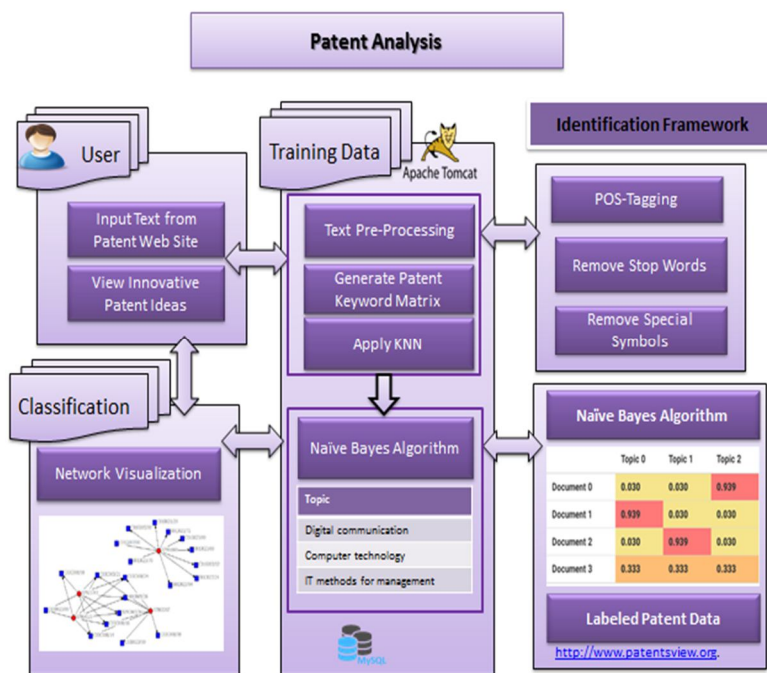


Figure: - System Architecture.

#### A. User Registration

At the time of the user registration, the user creates their id and password to our system.

#### B. Input Text from Patent Website

In this phase, the input text for the classification of patent data will be taken from the website. The input text can classify into Different category with the help of LDA algorithm.

#### C. Web-Site

The Data Set Website is <http://www.patentsview.org>. Data will be given as follows,

User can get text from this website and input this text into the system. Then the system will classify the text for Patent topic classification.

<b>Id</b>
<b>Type</b>
<b>Number</b>
<b>Country</b>
<b>Date</b>
<b>Abstract</b>
<b>Title</b>
<b>Kind</b>
<b>num_claims</b>
<b>Filename</b>

#### D. Text Preprocessing

The text preprocessing can be done by using natural language processing and information retrieval (IR).

Example: “Convolutional neural networks has been demonstrated high performance on image classification”

##### 1) Processing

a) The first stage if the preprocessing will be the POS Tagging. In POS Tagging the system processes a sequence of words, and attaches a part of speech tag to each word.

2) *Output*: “Convolutional/JJneural/JJnetworks/NNShas/VBZbeen/VBNdemonstrated/VBNhigh/JJperformance/NNon/INimage/N Nclassification/NN! /.”

a) Next we need to remove the stop words like and, is, where from the given data “Convolutional neural networks demonstrated high performance image classification!”

b) Now Remove the Special Symbols from the given sentence such as “! @#\$\$%^&\*()/\*-+”. “Convolutional neural networks demonstrated high performance image classification”

c) Now in last stage only pick the noun and adjectives from the remaining textual data “Convolutional neural networks high performance image classification” This output of stage four will be given it to the Next step to map keyword with Patent.

#### E. Generate a Patent-Keyword

From the output of the preprocessing we will calculate the Term Frequency of the each keyword Using TFIDF and group the keyword using the KNN Algorithm. A technical keyword Picked which reaches a high weight by having a high term frequency in its corresponding patent and a low document frequency in the entire patent collection.

#### F. Train LDA

For training, the data apply LDA algorithm. LDA helps for classification of the text in different Patent topic categories and assign them the weight.

#### G. Topic Modeling

Extracting semantic innovation Keywords and their relationships in group Using K-Nearest neighbor Algorithm.

#### H. Network Drawing

The output K-NN will be input for the network drawing and the multiple networks will be created.

#### I. Innovative Topics Categorization For Patents

1) *Using LDA*: The categorization is based on not only the body but also the header of the text or article. The metadata provide additional information that can be exploited and improve the categorization capability.

### IV. ALGORITHMS USED

#### A. Naive Bayes Classification Algorithm

Naïve Bayes Classifier is amongst the most popular learning method grouped by similarities that works on the popular Bayes Theorem of Probability- to build machine learning models particularly for disease prediction and document classification. It is a simple classification of words based on Bayes Probability Theorem for subjective analysis of content.

Use of Naïve Bayes Classifier:

1) If you have a moderate or large training data set.

2) If the instances have several attributes.

3) Given the classification parameter, attributes which describe the instances should be conditionally independent.

In Naive Bayes classifier (NB), it is assumed that a term’s occurrence is independent of the other terms. We want to find a class that gives the highest conditional probability given a document d:

$$\arg \max_{c \in C} P(c|d)$$

By Bayes rule [3],

$$P(c|d) = \frac{P(d|c) \cdot P(c)}{P(d)}$$

It is clear that

$$P(c) = \frac{|c|}{\sum_{c' \in C} |c'|}$$

And P (d) can be ignored since it is common to all classes.

There are two ways to compute P (d|c) based on the representation: either binary or term frequency-based. We show how to compute P (d|c) for the latter. Let Nit be the number of occurrences word wt in document di, and V the vocabulary size. Then P(di|c) is the multinomial distribution:

$$P(d_i|c) = P(|d_i|)|d_i|! \prod_{t=1}^{|V|} \frac{P(w_t|c)^{N_{it}}}{N_{it}!}$$

P (|di|)|di|! Is also common to all classes and thus can be dropped. Finally, the probability of word wt. in class c can be estimated from the training data:

$$P(w_t|c) = \frac{1 + \sum_{i=1}^{|D|} N_{it}P(c|d_i)}{|V| + \sum_{s=1}^{|V|} \sum_{i=1}^{|D|} N_{is}P(c|d_i)}$$

Where D is the training dataset.

### B. K-Nearest Neighbors Algorithm

The K-NN classifier is based on the assumption that the classification of an instance is most similar to the classification of other instances that are nearby in the vector space. Compared to other text categorization methods such as Bayesian classifier, K-NN does not rely on prior probabilities, and it is computationally efficient. The main computation is the sorting of training documents in order to find the k nearest neighbors for the test document.

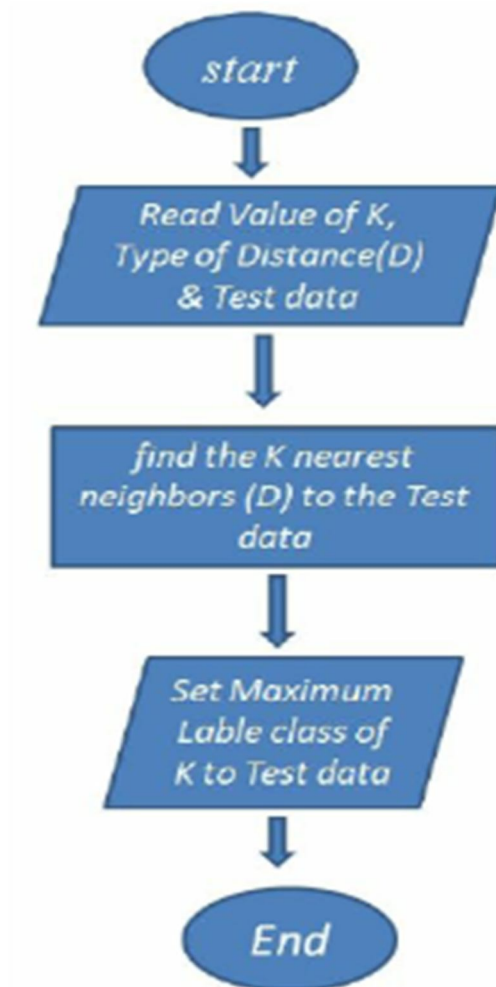


Figure: - Step wise KNN Algorithm

### V. RESULTS AND DISCUSSIONS

#### A. Registration Form

Above figure represents User Registration Form if the person is new user then he/she must compulsory fill the required details mentioned in the form and get register to the system.

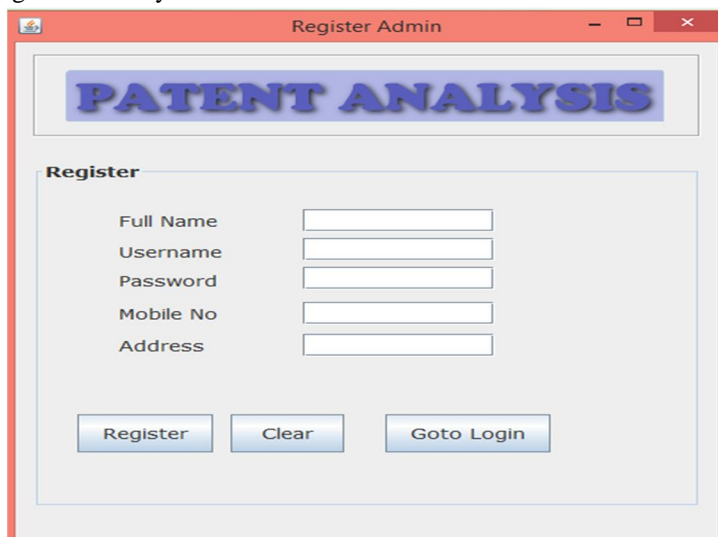


Figure: - User Registration Form

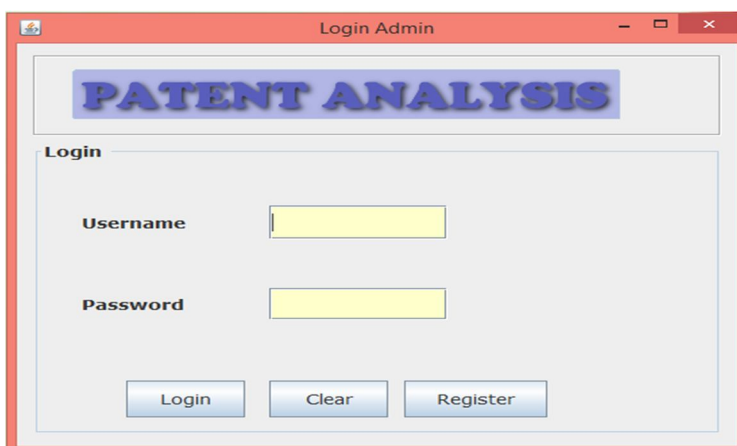


Figure: - User Login Form

Above figure represents User Login Form where user have to fill the required details like registered “username” and “password” for successful LOGIN.

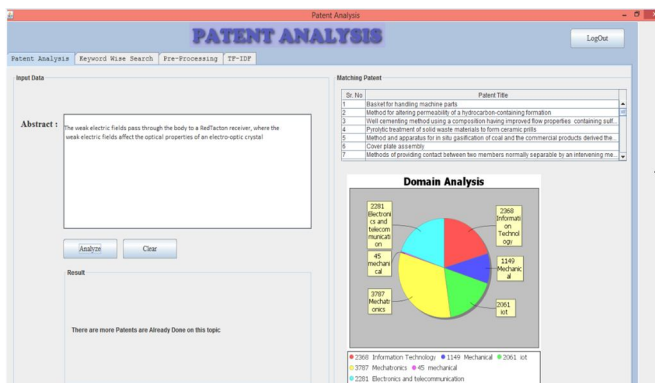


Figure: - Patent Analysis Form

Here in Input Data the user have to write the **Abstract** of the paper and after clicking on “**Analyze**” button Matching patents with its patent title are analyzed by the system and the Domain analysis pie chart is displayed with its results.

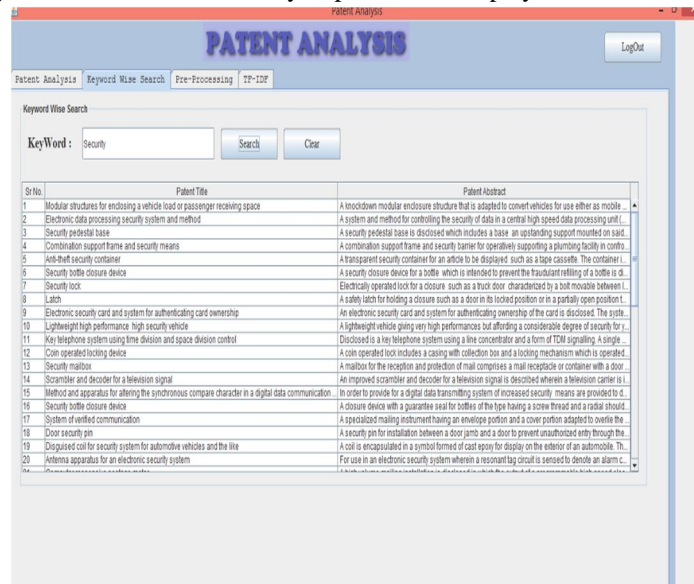


Figure: - Keyword Wise Search.

The Keyword Wise Searching mainly accepts input in the form of keywords and based on that the patent title as well as patent abstract is displayed.

1) *Preprocessing*: In data preprocessing following operations are performed;

- a. Remove Special Symbols
- b. Remove Stop Words
- c. Generate TF-IDF

The output of these operations is displayed in RESULT section

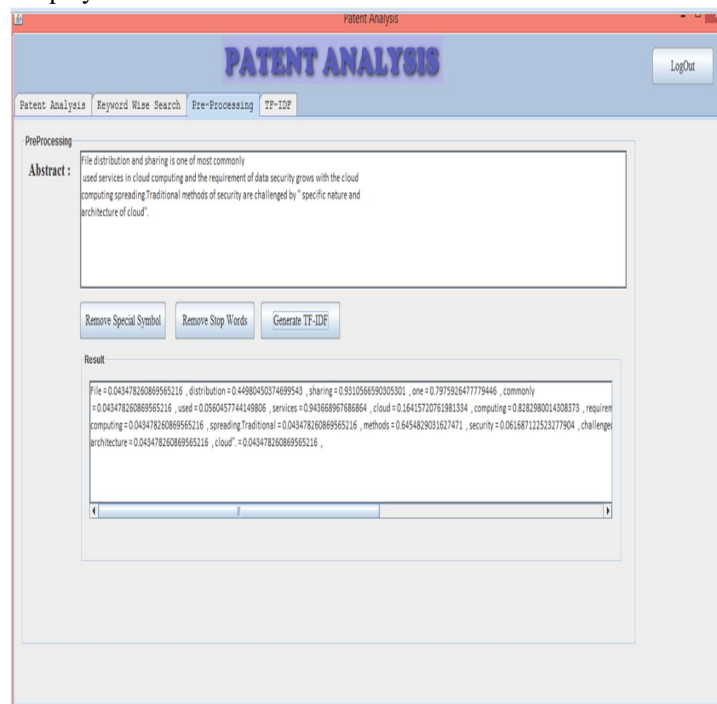
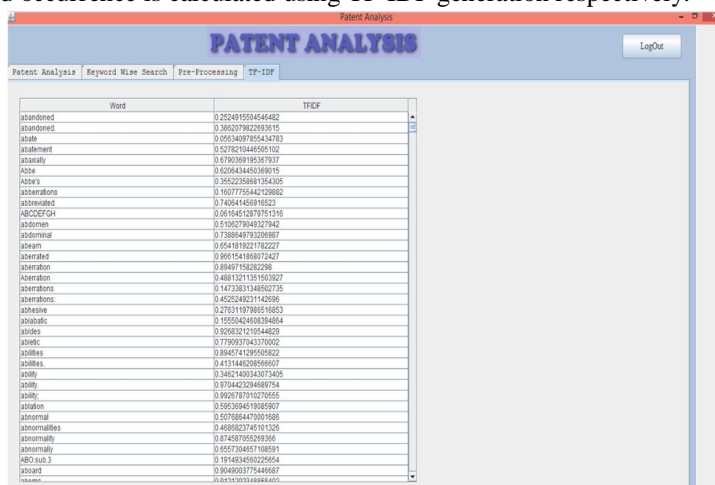


Figure: - Preprocessing.

The frequency of particular word occurrence is calculated using TF-IDF generation respectively.



Word	TFIDF
abandoned	0.2524915554545482
abandoned	0.389207782283915
abate	0.954349795424383
abatement	0.527821044505102
abatement	0.8790369195367937
abate	0.820943445050515
abate's	0.355225881154305
abaterations	0.160775542129882
abatemented	0.140444691623
ABCOEFGH	0.06164512879751316
abdomen	0.510627304327942
abdominal	0.7389848753209687
abdom	0.6541819221782227
aberrated	0.9651541898072427
aberration	0.88497156262296
aberration	0.48813211351953927
aberrations	0.14733831348502735
aberrations	0.462526823142896
abhesive	0.275119789515853
abhesive	0.1555042498334854
abhesive	0.90823211544489
abhesive	0.779093743370002
abhesive	0.884574129550522
abhesive	0.4171846238956607
abhesive	0.34821400343073405
abhesive	0.970427294689754
abhesive	0.9202781716278956
abhesive	0.5953548159085007
abnormal	0.5077884473019385
abnormalities	0.4698823748181285
abnormality	0.874587055289306
abnormality	0.657304857108591
ABCI sup 3	0.191484585225554
aband	0.904903175446587
aband	0.8154503342884405

Figure: - Generate TF-IDF

## VI. CONCLUSION

Patent analysis and patent technology indicators topics that is established to be one of the most essential and viable data sources, which should be a fundamental part in educational strategic like undergraduate (UG) as well as postgraduate (PG) and curriculums' plans. In this survey literature review of different author's in the area of patents analysis, patents technology indicators, and its applications are elaborated in this paper. Literature was able to identify many potential research gaps, which worth for further investigation.

## REFERENCES

- [1] D. Bonino, A. Ciamarella, and F. Corno. Review of the state-of-the-art in patent information and forthcoming evolutions in intelligent patent informatics. *World Patent Information*, 32(1):30 – 38, 2010.
- [2] Y. Yang, L. Akers, T. Klose, and C. B. Yang, "Text mining and visualization tools - Impressions of emerging capabilities," *World Patent Inform.*, vol. 30, no. 4, pp. 280 – 293, 2008.
- [3] M. Lupu, K. Mayer, J. Tait, and A. J. Trippe, Eds., *Current Challenges in Patent Information Retrieval*. Springer, 2011.
- [4] C. Chen, *Mapping Scientific Frontiers - The Quest for Knowledge Visualization*, 2nd Edition. Springer, 2013.
- [5] K. Börner, *Atlas of Science: Visualizing What We Know*. The MIT Press, 2010.
- [6] Alok Khode and Sagar Jambhorkar "A Literature Review on Patent Information Retrieval Techniques" *Indian Journal of Science and Technology*, Vol 10(37), DOI: 10.17485/ijst/2017/v10i37/116435, October 2017.
- [7] Longhui Zhang, Lei Li and Tao Li "Patent Mining: A Survey" *SIGKDD Explorations Volume 16, Issue 2*.
- [8] C. Lin. "The exploration of the relationship between the growth rate of Listed Companies' stock price and the specific patent indicators", *Research Portal, Science & Technology Policy Research and Information Center*, website: <https://portal.stpi.narl.org.tw/index/article/10113>.
- [9] M. Hirschey, V. J. Richardson and S. Scholz. "Value Relevance of Nonfinancial Information: The Case of Patent Data", *Review of Quantitative Finance and Accounting*, 17(3), 223-235, 2001.
- [10] R. Kapoor, M. Karvonen, S. Ranaei and T. Kässi." Patent portfolios of European wind industry: New insights using citation categories", *World Patent Information*, 41, 4–10, 2015.
- [11] G. J. Yu and K. H. Hong. "Patents and R&D expenditure in explaining stock price movements", *Finance Research Letters*, 19, 197–203, 2016.
- [12] F. Chan, Dora Marinova and ichael McAleera. "Rolling regressions and conditional correlations of foreign patents in the USA", *Environmental Modelling & Software*, 20(11), 1413–1422, 2005.
- [13] Paolo Federico, Florian Heimerl, Steffen Koch, and Silvia Miksch, "A Survey on Visual Approaches for Analyzing Scientific Literature and Patents" *IEEE* 2016.



10.22214/IJRASET



45.98



IMPACT FACTOR:  
7.129



IMPACT FACTOR:  
7.429



# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089  (24\*7 Support on Whatsapp)