Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X
IMPACT FACTOR: 6.017

IJCSMC, Vol. 6, Issue. 4, April 2017, pg.222 – 227

Patent Trend Analysis and Future Prediction

Dinesh Malpure, Yogesh Botre, Darshan Bhansali, Rohan Bhagi
K.K.Wagh Institute of Engineering Education & Research, Savitribai Phule Pune University, Nashik
dkmalpure@gmail.com, yogeshbotre123@gmail.com, dpbhansali@gmail.com, rohanbhg4@gmail.com
(+91) 73876316666, 9270706191, 9028796692, 7875257626

Abstract: Patents are a vital source of information for innovators, research developers, and policy makers of the technology domain. Information and meta-information related to these national and international patents is a critical asset for every innovative company. These patents are lengthy and are in technical terms, which require efforts for analysis. To address this issue, appropriate mining techniques are applied. These techniques involve knowledge discovery from available patent database allowing researchers to obtain an outline of the entire patent information, the algorithms used and the current trend analysis related to each domain. Also, the innovators or any other users are always eager to know which domains have future scope for registering patents, and from which domain the next big thing can come from. Thus, this paper aims at proposing a system representing a summary of patent information and innovative solutions in patent informatics, concerning the analysis, and also aims to provide a simple and suitable system for future trend prediction in various domains of technology.

Keywords– Information retrieval, patent mining, cluster generation, analysis, emerging trend prediction.

INTRODUCTION

Patent mining is extension of data mining as the tools are designed to handle unstructured or semi structured data sets. Data mining mainly focuses on analysing structured data. Current patent users are not only patent domain experts, but include new occasional actors such as managers, industrial researchers, academic faculty, and so on, each needing a different set of functionalities and a different degree of complexity. And users are always on the lookout for writing new patents under their name. Hence, special emphasis is given to the increasing variety of users that can benefit from easy access to patent information and future trend prediction, which can help users innovate and write new patents.

The increasing amount of patent applications and the growing need to access patent information make the task of patent analysis become vital to:

- Analyze large amounts of patent data that is expensive being done manually.
- Enhance the quality of generating useful information
- Support decision making processes to eventually improve the quality of the patents.
- Provide suggestions for further evolution in a particular domain of patents.
- Determine the patentability of their inventions.

- Avoid infringing other inventors' patents.
- Identify key trends in specific technical fields of public interest such as those relating to health or to the environment and provide a foundation for policy planning.

A typical patent specification has the following sections:

1. Title of invention: Title of inventions i used to educate the user with an initial description of the main content of the invention. Titles of inventions in patent documents form a vital part as they can be used for easy searching from a long list of patents. They are, as a rule, printed on the first page of patent documents and normally also included in the entries in Official Gazettes.

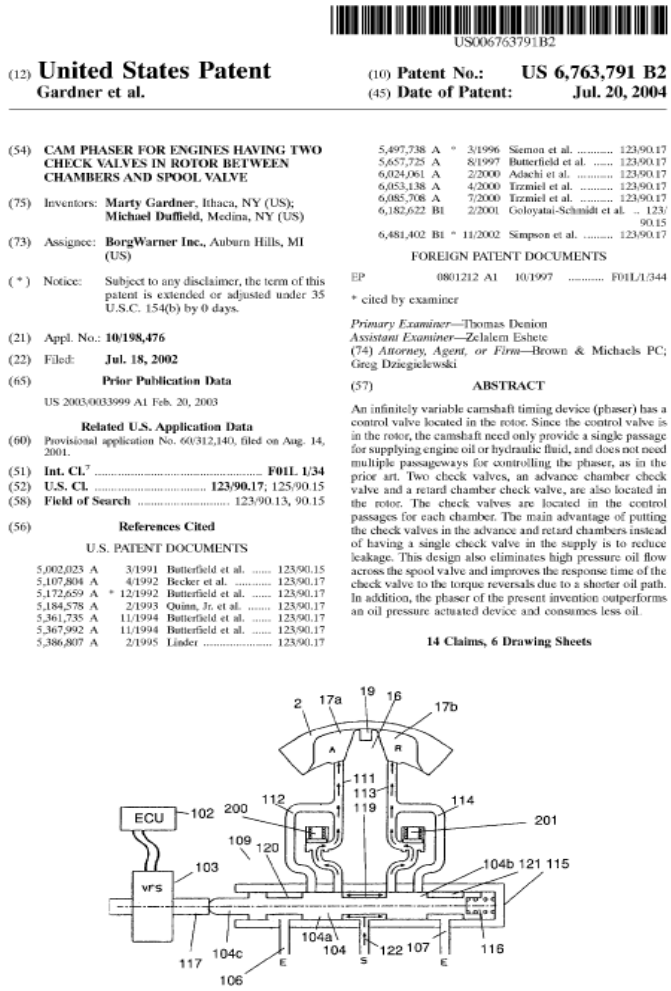
2. Applicant: This section contains the name or names of the applicants, applying for a certain patent.

3. Background: This section describes to what domain the invention belongs.

4. Description: This is an exact and simple explanation of the current problems, their current solutions and the advantages and disadvantages of these solutions. It also describes the new methods and technologies being used by the inventor with the help of examples.

5. Claims: Claims are the essence of a patent. The claims define the invention which the inventor holds as his exclusive property and has the right to exclude others from making, using and selling.

6. Citation and references: This contains references to the technology and methodology used in the invention, and are usually helpful for the patent examiner during the patent granting procedure.



US006763791B2

(12) **United States Patent**
Gardner et al.

(10) **Patent No.: US 6,763,791 B2**
(45) **Date of Patent: Jul. 20, 2004**

(54) **CAM PHASER FOR ENGINES HAVING TWO CHECK VALVES IN ROTOR BETWEEN CHAMBERS AND SPOOL VALVE**

(75) Inventors: **Marty Gardner**, Ithaca, NY (US); **Michael Duffield**, Medina, NY (US)

(73) Assignee: **BorgWarner Inc.**, Auburn Hills, MI (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 0 days.

(21) Appl. No.: **10/198,476**

(22) Filed: **Jul. 18, 2002**

(65) **Prior Publication Data**

US 2003/0033999 A1 Feb. 20, 2003

Related U.S. Application Data

(60) Provisional application No. 60/312,140, filed on Aug. 14, 2001.

(51) **Int. Cl.**[7] **F01L 1/34**
(52) **U.S. Cl.** **123/90.17**; 125/90.15
(58) **Field of Search** 123/90.13, 90.15

(56) **References Cited**

U.S. PATENT DOCUMENTS

5,002,023 A 3/1991 Butterfield et al. 123/90.15
5,107,804 A 4/1992 Becker et al. 123/90.17
5,172,659 A * 12/1992 Butterfield et al. 123/90.17
5,184,578 A 2/1993 Quinn, Jr. et al. 123/90.17
5,361,735 A 11/1994 Butterfield et al. 123/90.17
5,367,992 A 11/1994 Butterfield et al. 123/90.17
5,386,807 A 2/1995 Linder 123/90.17
5,497,738 A * 3/1996 Siemon et al. 123/90.17
5,657,725 A 8/1997 Butterfield et al. 123/90.17
6,024,061 A 2/2000 Adachi et al. 123/90.17
6,053,138 A 4/2000 Trzmiel et al. 123/90.17
6,085,708 A 7/2000 Trzmiel et al. 123/90.17
6,182,622 B1 2/2001 Goloyatai-Schmidt et al. .. 123/90.15
6,481,402 B1 * 11/2002 Simpson et al. 123/90.17

FOREIGN PATENT DOCUMENTS

EP 0801212 A1 10/1997 ..... F01L/1/344

* cited by examiner

Primary Examiner—Thomas Denion
Assistant Examiner—Zelalem Eshete
(74) *Attorney, Agent, or Firm*—Brown & Michaels PC; Greg Dziegielewski

(57) **ABSTRACT**

An infinitely variable camshaft timing device (phaser) has a control valve located in the rotor. Since the control valve is in the rotor, the camshaft need only provide a single passage for supplying engine oil or hydraulic fluid, and does not need multiple passageways for controlling the phaser, as in the prior art. Two check valves, an advance chamber check valve and a retard chamber check valve, are also located in the rotor. The check valves are located in the control passages for each chamber. The main advantage of putting the check valves in the advance and retard chambers instead of having a single check valve in the supply is to reduce leakage. This design also eliminates high pressure oil flow across the spool valve and improves the response time of the check valve to the torque reversals due to a shorter oil path. In addition, the phaser of the present invention outperforms an oil pressure actuated device and consumes less oil.

14 Claims, 6 Drawing Sheets

RELATED WORK

The evolution of patent mining and patent analysis is done by bibliometric and keyword based network analysis. Cluster analysis technique is applied to carve out the evolution pattern from the papers, based on the keywords that have been mined from the abstracts of papers. These keywords are also useful for searching of relevant papers as mining is also a vital part of the patent analysis procedure.

The evolution of patents is in two main stages:

1) Patent analysis evolution and 2) Patent mining evolution. Various measures and statistical indicators are used for the analysis of such bibliometric data and only then are the network and cluster analysis techniques applied for extracting patterns which are more complex to understand and difficult to cite [1].

Patent mining is a technique that is vital for the searching of cutting-edge technology that would be beneficial for the given production and its development [2] .Text mining is another useful method, mainly advantageous for the extraction of useful knowledge from a large bank of unstructured patent data. This unstructured data can be in the form of images, HTML data that can be converted into a structured form with the help of algorithms and text mining techniques such as Natural Language Processing, Semantic Analysis. Another approach is Virtualization, which makes use of the citation method to establish relations between patents and the trends. Patent search is beneficial if someone has a patent related to a specific idea or an object. Large size of the patent documents call for an efficient parsing and faster retrieval mechanism. There is a need for an indexing system which will be useful for the search and retrieval of patents.

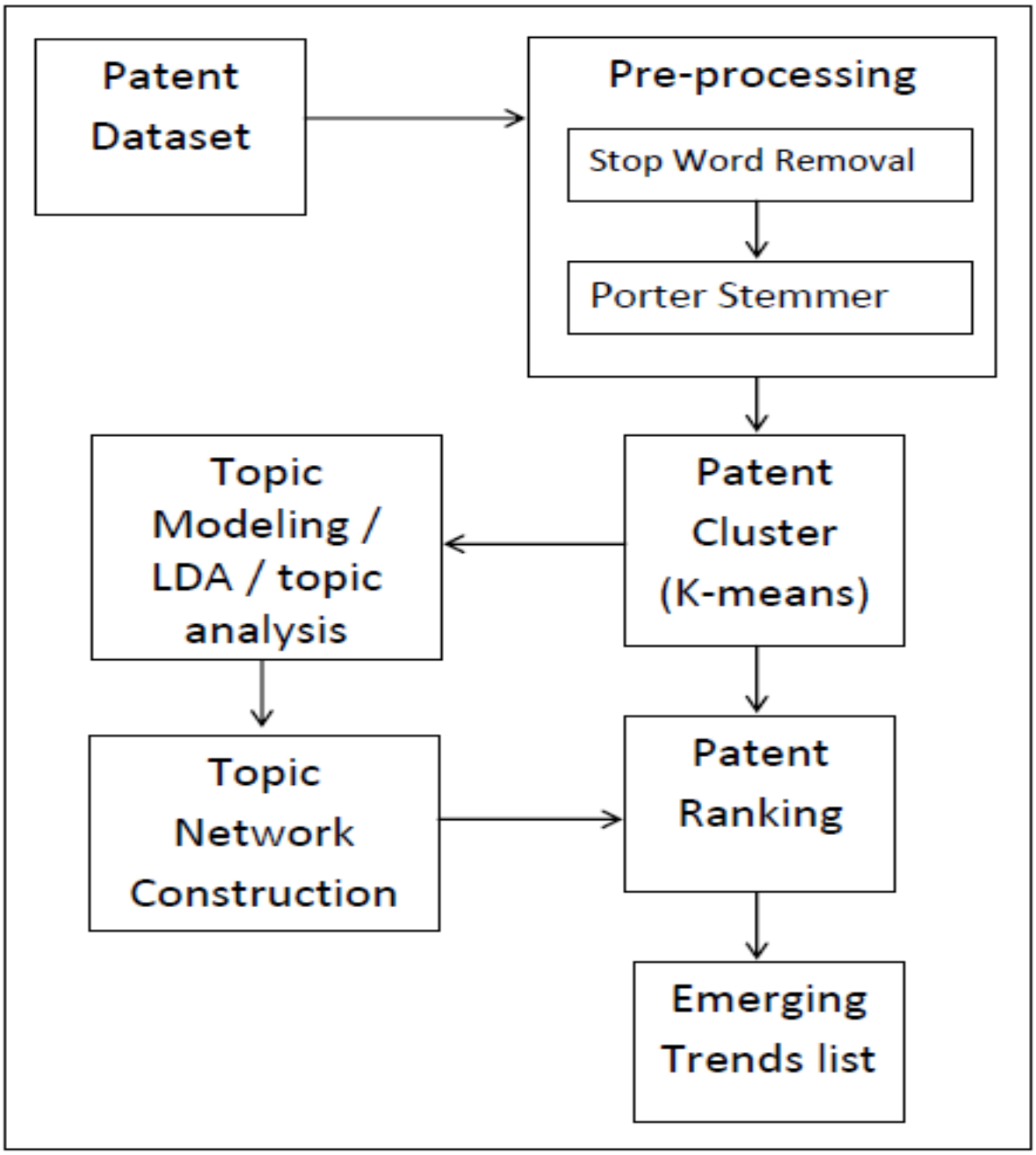
Factors that define the value of the patent search system are: (a) Coverage: which databases are included, from which domains, countries, or organizations and how this system is updated. (b) Discovery & Ranking: To what level of accuracy does the system discover the relevant patents according to the query and how does it rank these patents.

The main features are: (1) Topic Generation (2) Patent Storage (3) Information Retrieval and (4) Trend Analysis Technology Forecasting.[3]

The increasing amount of patent applications and the necessity to access this patent information proves the task of patent analysis vital for 1) analysis of the large amount of patent data which is not affordable for humans, to do on their own, 2) enhancement in the quality of the information generated and 3)support decision making processes to eventually enhance the quality and efficiency of the patents. Patent intelligence is used to encourage the development of innovative products, devise technology strategies, and reveal legal/business insights amid the technical transformation. To obtain semantically meaningful topics of interest, the TF*PDF algorithm is used, which allows for the generation of hot topics over time[4].The rights granted to applicants by patents are an important method to stop others from making, using, selling, or importing an invention illegally. It is important for assignees and attorneys to fully understand the patents in a legal situation. For example, assignees can inspire their RD direction and industrial solutions, reveal business trends or help decisions by reading the patent documents precisely in the field where they are located. For discovering rare features from massive patent documents, a new indicator is used to measure the rareness of features with low frequency. The method applies term frequency (TF) and inverse document frequency (IDF) as an apriori weight to the measurement of max confidence. With the experimental results based on a real patent data set, the method is proven to be effective in finding rare features in patent documents[5]. Pre-filling and post-filling techniques are performed during patent analysis. Pre-filling process includes two major tasks classification and search of all relevant patent documents from patent databases and non-patent documents from online sources. In the process of post-filling, patent visualization patent valuation and Cross-language mining are performed. Various challenges in this domain include Figure-Based Patent Search, Product-Based Patent Search, Large-Scale Patent Retrieval Multi-Label Hierarchical Patent Classification and Cross-Domain Patent Recommendation [6].

SYSTEM ARHCITECTURE

This paper describes the procedure for patent analysis and future trend prediction. First, we gather the patent files in xml format for the generation of a database. These documents cannot be utilized for mining purposes until they are pre-processed. Next, data pre-processing methods are applied, which include stop word removal, document term matrix generation followed by feature vector generation. After we extract the thematic words, clustering is applied to form appropriate clusters. Then, we filter out various topics that have appeared frequently and then rank them based on frequency. And in the end, the emerging trends are predicted.



***A. Data Base*:** Patent files are downloaded in XML format. The parsing is first done to abstract the details such as Patent ID, Invention title, Abstract, Summary, Claims using Java by removing the HTML tags. Java code is iterated to extract all the metadata from a given patent file. Database comprising of information about patents is obtained.

***B. Data Pre-processing*:** The data pre-processing step consists of: stop word removal, feature vector generation.

1.Stop word removal: This step can be considered as the most vital part of the data pre-processing step. This method removes unnecessary words from the patent and generates a set of unique words from each patent. The accuracy of this step affects the final results as only after this, can we calculate words' frequency. Porter stemmer algorithm is applied to reduce a given word to its relevant root word. This enhances the probability of occurrence of a word in multiple patents, thereby improving the precision and performance of clustering and analysis.

2. Document term matrix generation: This denotes the frequency of occurrence of each term appearing in a collection of document. In document term matrix, columns are used to represent documents in the collection and row corresponds to terms. **Term frequency–inverse document frequency**, is used to retrieve information about a numerical statistic that reflect how important a word is to a document in a collection. In the case of the **term frequency** tf(*t*,*d*), the simplest choice is to use the *raw frequency* of a term in a document, i.e. the number of

times that term *t* occurs in document *d*. If we denote the raw frequency of *t* by $f_{t,d}$, then the simple tf scheme is $\text{tf}(t,d) = f_{t,d}$.

3. Feature vector generation: A **feature vector** is an n-dimensional vector of numerical features that represents some object. A feature vector for each document term is calculated. This feature vector serves as a foundation for the clustering process.

C. Clustering*:*

K-means clustering: It is one of the most frequently used unsupervised algorithms for clustering. K-means clustering is exploratory data analysis technique. This is non-hierarchical method of grouping objects together. Cluster analysis or clustering is the task of grouping a set of objects in such a way that objects in same group(called a cluster) are more similar to each other than to those in other groups(clusters).

Step 1: Input:

a) Feature Vector generated in pre-processing process.

b) Number of clusters.

Step 2: Initialize the cluster centroid.

Step 3: Calculate cosine similarity.

Cosine similarity: Cosine similarity can be computed amongst vectors. It is a *similarity measure that* gives a measure of how similar two documents are likely to be. Given two vectors of attributes, *A* and *B*, the cosine similarity, *cos(θ)*, is represented using a dot product and magnitude as

$$similarity = \cos(\theta) = \frac{A \cdot B}{\|A\|\|B\|} = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2}\sqrt{\sum_{i=1}^{n} B_i^2}}$$

where A_i and B_i are components of vector A and B respectively. In the case of information retrieval, the cosine similarity of two documents will range from 0 to 1 , as the term frequencies (tf-idf weights) cannot be negative.

Step 4: Move on to next observation and calculate cosine similarity.

Step 5: Calculate cosine similarity for the next observation, assign next observation based on maximum cosine similarity value and update the cluster centroids.

Step 6: repeat the above procedure for stable Clusters.

D. Topic modelling LDA*:* Consider a set of documents and some fixed number of K topics to discover, and want to use LDA to learn the topic representation of each document and the words associated to each topic.Go through each document, and randomly assign each word in the document to one of the K topics.

Notice that this random assignment already gives you both topic representations of all the documents and word distributions of all the topics (albeit not very good ones).

So to improve on them, for each document d.

Go through each word w in d…

And for each topic t, compute two things:

1) p(topic t | document d) = the proportion of words in document d that are currently assigned to topic t

2) p(word w | topic t) = the proportion of assignments to topic t over all documents that come from this word w.

Reassign w a new topic, where we choose topic t with probability p(topic t | document d) * p(word w | topic t).

Repeat above steps until steady state which give the keywords of different number of topics which was given to LDA.

E. Topic Network Construction: LDA give the different topic keywords. In this step network is generated between different topic.

F. Patent Ranking: Topic network gives different topic and their interconnection from this information patents are ranked on the basis of most frequent used topic ranking of patents are used to find emerging trends from different patents.

In our system, we are finding emerging trends in every domain by considering all domains and citations between these domains.

User will enter the domain name, then our system gives the list of all emerging trends and technologies in the form of keywords.

CONCLUSION

This paper proposes a Patent Prediction system that can be used to provide the scope of particular domain to the users by complete analysis of the patent. The goals and objective are to achieve a holistic patent predicted output facility for the user's input query request. A brief description of the project is given by defining the problem statement and project idea. Also, the motivation behind this project is that this will help non-technical users or first time users get a better understanding of the patent scenario, be it filing for a patent or searching for one and users can get a detailed and quick review of the existing advancements in required technological areas, and, also have an idea as to where and in which technological area or domain lies the most scope for future patent filing. The hardware and software requirements, outcome and applications of the project have been stated. Time and cost estimation has been performed. Risk identification, analysis and management related to the project has also been done. A project schedule has been designed which contains major tasks and a timeline chart. Team structure and tasks of members are also defined. Architectural design and UML diagrams have been made relating to the problem statement.

REFERENCES

[1] N. Mattas, Smarika, D. Mehrotra ,"Comparing data mining techniques for mining patents", Fifth International Conference on Advanced Computing & Communication Technologies, 2015.
[2] A.Supraja, S.Archana, S.Suvetha, "Patent Search and Trend Analysis", IEEE International Advance Computing Conference (IACC), 2015.
[3] K. Ly, N. Shin, S.J. Yoo," Hot Topic Detection and Technology Trend Tracking for Patents utilizing Term Frequency and Proportional Document Frequency and Semantic Information", 978-1-4673-8796-5/16/$31.00 2016 IEEE, 2016.
[4] M. Guo, H. Yuan, Y. Qian, "New Method for Rare Feature Extraction in Patent Documents", 978-1-50902842-9/16/$31.00 ©Q016 IEEE, 2016.
[5] L. Zhang, L. Li, T. Li," Patent Mining: A Survey", ex SIGKDD Explorations Volume 16, Issue 2, 2013.
[6] Zalányi,Kinga Makovi,, Zoltán Somogyvári,Katherine "Prediction of Emerging Technologies Based on Analysis of the U.S. Patent Citation Network, 2012".
[7] Patent Maintenance Recommendation with Patent Information Network Model, 2011.
[8] Two simple predictions algorithms to faciliate text production, 2015.
[9] Patent Search and Trend Analysis.

