IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 1, October 2022
Impact Factor: 6.252

Need for 'Patent Search' Modernisation and Potential for its Acceleration Leveraging AI/ML Models

Manish R. Potdar[1] and Dr. B. J. Mohite[2]
Principal Director, L&T InfoTech Limited, Pune, India[1]
Assistant Professor, ZEAL Institute of Business Administration, Computer Application & Research, Pune, India[2]
manishrpotdar@gmail.com[1] and babasaheb.mohite@zealeducation.com[2]

Abstract: *Intellectual Property Rights (IPR) have played a crucial role in promoting innovations across the globe and of late the industry has seen a steep rise in innovation activity. There is an unprecedented urgency to help global IP offices in shortening the processing time, managing bureaucratic delays, and improving operational transparency.*

Patents are one of the most critical IP types among the 6 types where there is significant traction over the last couple of years. From ideation to the grant of the patent and its commercial use, the past patent data is required to be searched for and referred to for; 1] patentability assessment, 2] invalidity assessment, etc. Therefore, 'Search' becomes the most critical process across the patent lifecycle.

Literature study indicates that issues involved in patent search (when traditional search techniques are employed) usually are around 1] data processing errors, 2] errors due to language pitfalls, 3] errors due to faulty syntax, and 4] classification errors. These erroneous searches result in a large number of false positives and false negatives.

Artificial Intelligence (AI) and Machine Learning (ML) are leading the wave of technology development - both from a research and development perspective as well as their commercial use. Adopting these next-generation technologies presents great potential to help address the growing challenges in the patent search process.

AI/ML based models are suitable predominantly for multi-lingual search, handling diverse data formats, image comparisons, and keyword matching. As IP databases across countries still lack standardization, advanced technologies such as generative AI are best suited to help accelerate the patent search process.

Feasibility assessment of leveraging various AI/ML models to address efficiency and effectiveness issues of patent search can be performed through a 3-part framework (3i) focussing on various dimensions such as Integrate, Infer, and Intelligence.

AI/ML model applicability can be assessed against specific objectives of each part viz.

*1.**Integrate** – integration with various patent databases,*

*2.**Infer** – data extraction and transformation into a standardized data set suitable for comparative analysis and*

*3.**Intelligence** – comparison, matching, and decision-making for search objectives.*

At a global scale, further deliberations and studies on this subject are of immense value in the areas of knowledge and policy-making thereby benefiting practitioners, the academic fraternity, and society.

Keywords: Intellectual Property, Patent, Patent Search, AI/ML Models, Intelligence

I. INTRODUCTION

Intellectual Property Rights (IPR) have played a crucial role in promoting innovations across the globe and of late there has been a steep rise in innovation activity. The ongoing crisis (Covid-19) has impacted the Intellectual Property industry both positively by creating a need for accelerated commercialization of innovations and negatively by creating a funding challenge on the continuity of innovation efforts. This dual impact has generated an unprecedented urgency to help global IP offices in shortening processing time, manage bureaucratic delays, and improve operational transparency. Technology advancements across the globe have been instrumental in the economic development and well-being of human beings. Huge efforts have been invested across the globe in researching and developing new technologies to help

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 1, October 2022
Impact Factor: 6.252

improve effectiveness and efficiency in human life. Artificial Intelligence (AI) and Machine Learning (ML) continue to lead the wave of technology development - both from a research and development perspective as well as commercial use. AI/ML-based models have a wide range of implementation opportunities e.g., automating repeated/mechanical decision-making, pattern matching, and generating the resulting image, etc.

AI/ML technology presents great potential to help address the growing challenges of IPR. The subject has great relevance in the context of India with its initiatives including "Creative India: Innovative India" as the Indian IP industry continues to be immature and constrained by limited investments in modernising the IP lifecycle.

This conceptual research paper examines the challenges faced by the IP industry (particularly patents) during the IP assignment lifecycle. The paper also presents a skeletal framework for leveraging AI/ML models in the 'Patent Search' process along with the recommended areas of further research.

II. LITERATURE REVIEW

2.1 What is an Intellectual Property (IP)?

According to World Intellectual Property Organisation (WIPO), Intellectual property (IP) is inventions (human mind creation); scholarly and artistic works; designs and symbols, names and images used in commerce. IP is protected in law by patents, copyright, trademarks, etc. enabling people to earn recognition or financial benefit from what they invent or create.

There are 6 types of IPs viz.

1. Copyright – It is a legal term used to describe the rights that creators have over their literary and artistic work,
2. Patent – It is an exclusive right granted for an invention,
3. Trademark – It is a sign capable of distinguishing the goods or services of one enterprise from those of other enterprises,
4. Industrial Design - It constitutes the ornamental or aesthetic aspect of an article,
5. Geographical Indication - Geographical indications and appellations of origin are signs used on goods that have a specific geographical origin and possess qualities, a reputation, or characteristics that are essentially attributable to that place of origin and
6. Trade Secrets – These are IP rights on confidential information which may be sold or licensed.

Particularly, patents in many countries are an essential form of competitive advantage. It is an incentive for companies to invest as they gain competitive as well as commercial advantage. It is also a medium to incentivise economically efficient research and development.

2.2 Importance of Patent Search

Innovators and companies across the globe invest considerable money and effort in protecting their inventions/ideas for goodwill, commercial benefits, individual identity, etc. From ideation to the grant of the patent and its commercial use, there is a need to search for existing innovations and associated patent details across multiple stages. Therefore, from an efforts standpoint, patent search becomes a critical process across the patent's assignment lifecycle.

Following are the various scenarios where the patent search is carried out –

- Patentability – This is a type of search carried out to assess the originality of the invention before filing the patent.
- Right to operate - This is a type of due diligence for finding any active patent or pending patent applications impacting freedom to operate.
- Invalidity – This type of search is conducted for various reasons such as supporting legal instances, acquiring rights, competitive analysis on innovations, etc.
- State of art - This type of search is conducted to gain an overall understanding of the patent field in a specific technology or industry domain.

Over the years, the IP industry lacked significant investment and continued to be an underdeveloped capability. There are various challenges faced by the industry including political, technological, commercial, skill-related challenges, etc. With increasing innovation activity across the globe, there is a need to modernize the overall IP industry process to bring efficiency and cost-effectiveness.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 1, October 2022
Impact Factor: 6.252

2.3 Artificial Intelligence leading the wave of Next-Generation Technologies

Fundamentally, Artificial Intelligence (AI) is a technique that enables machines to mimic human behavior. AI model is a tool or an algorithm which, based on the input data set, can derive decisions without human intervention. The input data set is utilized to identify patterns in the data which enable decision-making in a certain typical scenario. AI models are particularly suitable for solving complex problems while providing higher efficiency, cost savings, and accuracy as compared to simple methods.

Machine Learning (ML) is a subset of AI techniques that leverage statistical methods for machines to improve with experience. ML models are used in most aspects of human life and businesses with the purpose of intelligent decision-making, automation, and life-cycle acceleration. These models enable software to make accurate predictions based on the understanding of patterns by screening through a large volume of data.

Supervised ML models are trained over data labeled with input and output. These models require specific instructions for evaluations, pattern identification, and predictions. On the other hand, unsupervised ML models work on unlabelled data. The patterns are identified through continued consistency. Models implementing a combination of both supervised and unsupervised models are termed, Semi-supervised ML models.

2.4 Different Types of AI models in the Industry

Following are a few AI models in the industry built with a set of objectives. This list is not a comprehensive one but gives an overall understanding of the development in the space.

Table 1: Types of AI Models

#	AI Model Type	Problem Solving Strategy	Usage	Prima-facie Applicability
1	Linear Regression	Find relationships between the input and output variables	Across domain-intensive industries – banking, insurance, etc.	May Not be Suitable
2	Deep Neural Networks	Artificial Neural Network (ANN) with multiple (hidden) layers between the input and output layers	Speech recognition, image recognition, and natural language processing (NLP)	Suitable
3	Logistic Regression	A statistical model that can predict the class of the dependent variable from the set of given independent variables	Solving binary classification problems	May Not be Suitable
4	Decision Trees	Data is divided into smaller portions that resemble the structure of a tree	Regression and classification problems	May Not be Suitable
5	Linear Discriminant Analysis	This is usually used when two or more classes are to be separated in the output	Various tasks in the field of computer vision, medicine, etc.	May be Suitable
6	Naive Bayes	Works on the assumption that the occurrence of any feature does not depend on the occurrence of any other feature	Used for both binary and multiple-class classifications. (Medical data classification, spam filtering, etc.)	May be Suitable
7	Support Vector Machines	A quick and efficient model that excels in analysing limited amounts of data	Applicable in binary classification, outlier detection, and regression problems.	May be Suitable
8	Learning Vector Quantization	Type of Artificial Neural Network that works on the winner-takes-all principle. It processes information by preparing a set of codebook vectors that are then used to classify other unseen vectors	Solving multi-class classification problems	May be Suitable

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 1, October 2022
Impact Factor: 6.252

9	K-nearest Neighbors	This algorithm works on the assumption that similar things (data) exist near each other	Simple supervised ML model used for solving both regression and classification problems	May not be Suitable
10	Random Forest	Uses multiple decision trees and makes the final prediction using the bagging method	Ensemble learning model useful for solving both regression and classification problems	May Not be Suitable

* Initial applicability analysis, needs further investigation

2.5 AI technology can play a role in IP Modernisation

Modernising with the use of AI techniques (such as hotspot and impact identification with correlation and pattern matching etc.) can help patent life-cycle by reducing elapsed duration, deriving optimal search volume, deriving geographical as well as technological breadth, and supporting predictive documentary reviews.

AI/ML models are found to be suitable predominantly in multi-lingual search, handling diverse data formats, image comparisons, and keyword matching. Because the IP databases across countries still lack standardization, advanced concepts such as generative AI will find more suitability in contributing to the acceleration.

With the help of AI-based semantic search engines and emphasizing the idea and key concepts, improved outputs can be extracted. AI/ML models prove to be advantageous for conducting a patent search as it decreases the manual workload, increases the accuracy and efficiency of the search results, and save plenty of time by considering the researcher's intent.

The literature review indicates that there has been an increasing interest amongst researchers in studying the potential use of next-generation technologies across IP activities. The research focus also spans deep research problems such as tokenisation and notarising of IP assets, drawing inferences in IP infringement, etc.

III. NEED AND SCOPE

3.1 Challenges faced in Patent Search

A patent search is a highly complex and specialised process. Literature study indicates that issues involved in patent search (when traditional prior art search techniques are employed) usually are around resulting in a large number of false positives and false negatives.

- Data processing errors,
- Errors due to language pitfalls,
- Errors due to faulty syntax,
- Classification error

Keyword-based search strategies are widely used by patent attorneys for prior art searches. Many times, this results in an inaccurate output considering a large volume of false positives and negatives which causes an additional burden of exclusion of documentation or erroneous granting of the patent. There are a set of common challenges the patent search process faces irrespective of the industry, type, and originating agency-innovator or outsourced agency of the IP.

These challenges include

Size and scale of the patent data – With the increase in awareness around protecting innovation rights, and forces like the covid-19 pandemic, there is an increase in patent filing in recent years. For multi-geography patents, the examination for originality turns out to be a too laborious and expensive process.

- Multi-lingual patents/databases – Not all patents are filed in the English language, especially in the countries such as China, Japan, and Korea where counterpart family members do not exist in English. There is a margin of error to be factored in using translation software
- IP database standards – Across geographies, there is still a lack of standard adoption when it comes to IP protection. Recently there have been efforts globally to collaborate across IP offices for developing a standard however earlier documentation continues to be in varying formats.
- Articulation/Terminologies – Major one, is inconsistency as either Dictionary synonyms were used, or Abstract

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 1, October 2022
Impact Factor: 6.252

and Creative Ways were used to describe the same very technology across patents. Thus, to conduct an exhaustive search, we need to consider these inconsistencies in the terminology across patents that deal with the same very technology.

Other general challenges faced during the patent search include –

- Non-availability of freely accessible patents and/or pending patent applications
- A very large volume of already published documents
- Complete manual search & analysis resulting in a higher cost of review.
- Non-patent literature databases (IEEE, Science Direct, Springer Link, etc.) are being paid, and the research papers are locked
- Dispersed and paid patent search platforms (XLSCOUT, Derwent Innovation, Questel Orbit, PatSnap, PatBase, and so on.)

3.2 Prioritizing challenges faced by Patent Search

Upon analysis, it is evident that there are 3 dominant categories of the challenges faced by the patent search – 1] System bureaucracy, 2] Volume of data and 3] discrete data sets. While bureaucracy needs an independent strategy to handle, the other two would primarily be focused on the efficiency and effectiveness of the search process and tools leveraged. The use of the next generation technologies will help in improving efficiency and effectiveness and can set up a strong foundation for long-term sustainable solutions.

The below conceptual framework and recommendations are scoped for efficiency and effectiveness of patent search with the influence of technology.

IV. RESEARCH METHODOLOGY

4.1 Patent Search - Solutions adopted by the industry so-far

Following are a few focused search strategies adopted by various entities while conducting the patent search –

A. Refined Search Methodology

- Refinement to the search methodology by [i] adopting a glossary of terms in the area of the innovation and area of application and [ii] bringing variations in search terms, synonyms, acronyms, etc. has helped make search efficient and effective.
- Enriched search by bringing in specially crafted logic individually for background titles, abstracts, descriptions, claims, etc.
- Use of different classification schemes.
- Perform extensive citation analysis; name-based analysis and file wrapper analysis for the patents
- Use of iterative process to improve the quality of search.

B. Use of Tools

- Aggregator tools ensure a majority of the global coverage while analysing the novelty. While doing so, multi-lingual non-English databases and geographies are also covered.

These measures turned out to be inadequate as they were laborious, manual, and suitable at the macro level. They were ineffective in conducting deeper analysis.

4.2 Conceptual framework of leveraging AI models for patent search

The objective of conceptualizing a new model is to

- Identify information that is necessary for AI processing of patents and related documents,
- Potentially bifurcate steps involved in patent search and group them logically into a structural form so that the application of various AI algorithms can be evaluated.

This approach will help increase the productivity and accuracy of decision-makers in the patent field.

The conceptual framework recommended is titled as ***3i* Model** and is proposed to be comprised of 3 parts viz. 1] Integrate, 2] Infer and 3] Intelligence. Each part performs a specific objective with the potential of leveraging AI

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 1, October 2022
Impact Factor: 6.252

algorithms. These parts logically group the activities performed during the patent search and critically evaluate the possibility of leveraging model capabilities.

The following diagram is a schematic representation of the ***3i* model** depicting 3 parts and groping of activities -

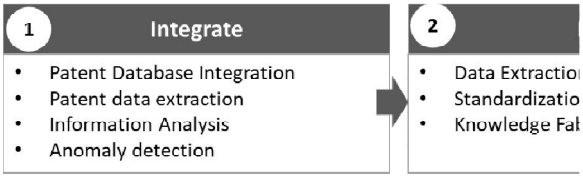

Figure 1: 3i Model – Conceptual View

***3i* Model** is proposed to be a set of algorithms that will perform multiple operations viz.

1. Extract relevant patent information from various databases,
2. Analyse the extracted information against the potential innovation under consideration for patent assignment
3. Generate intelligence for the decision thereby helping in the search process efficiency and effectiveness.

Following table details out objectives, functions, and output of each part along with examples

Table 2: Parts of Conceptual Model (3i Model)

Sr	Title	Functions	Examples	Outcome	AI/ML Model Applicability
1	Integrate: The objective is to integrate with various data sources & fetch data on a common platform	• Connect with various Patent databases • Extract patent data (diversified formats across databases) • Analyse data elements for various header information • Identify Anomalies	USPOTO, PATENTSCOPE, ESPACENET, Google Patents etc.	Extraction of patent data across databases	Classification & Regression, Outlier Detection
2	Infer: The objective is to extract critical information and standardize the same across data sources & data element	• Extract patent data into a commonplace • Patent data transformation for standardization across databases • Knowledge fabric comprised of patent information repository with standardized information	Patent specifications – Title, Field, Objective, Summary, Claims, Abstracts, etc.	Universe of patent data with relevant information in a standardized format	Classification & Regression, Translation, Recognition (Image),
3	Intelligence: The objective is to match and compare the data elements against pre- defined rules and identify matching parts	• Definition of various rules for comparison patent data • Actual matching and comparison of knowledge fabric elements with innovation under consideration for duplication, similarities, etc. • Decision and recommendations	Patentability, Freedom-to-Operate, The invalidity, State of the Art	Occurrences of similar innovation in the existing patent Guidance on the originality of the innovation	Matching, Filtering, Input/output Relationship

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 1, October 2022
Impact Factor: 6.252

4.3 Key Considerations for setting-up up the 3i Model Algorithms

To implement the proposed conceptual framework, the following are the key considerations that need detailed analysis and assessment for further development –

- Patent specification extraction
- Various types of patents and the key information captured during patent registration
- Correlation between various patent-related data points
- Various ways in which a patent can be analyzed for
- Boundary criteria that define the anomaly
- Key information and parameters for comparative assessment
- What defines false positive and false negative
- Different outcomes possible out of the comparative studies
- What are the decisions made and how are they currently made?

The analysis needs to be inclusive to include involved parties such as government organisations, legal representatives, innovators, private agencies, academia representatives, etc.

V. CONCLUSION

With increasing innovation activities across the globe, there is a significant increase in the patent filing. Across the lifecycle, patent search is an important and time-consuming activity. There is an opportunity to invest in technology adoption in resolving patent search-related challenges.

AI/ML-based algorithms have got the potential in modernising the patent search process thereby enabling efficiency and effectiveness in the process. To identify AI/ML-based algorithm adoption opportunities, a 3-part framework (3i Model) is visualized covering various dimensions such as integration with databases, creation of the standardized dataset for comparison, actual comparison, and intelligence-based decision making.

Each part of the conceptualized framework signifies a specific set of functions supported by peculiar characteristics. These characteristics pose peculiar challenges and present an opportunity to implement AI/ML-based algorithms.

5.1 Recommendations for Further Studies

It is recommended that a detailed evaluation of various AI/ML models should be done for their specific usage in accelerating the patent search. It is also recommended that a model should be developed considering the applicability and objective of the search process.

Designing AI algorithms tailored to patent-related tasks requires expertise in both AI technology and patents. To identify independent areas for further studies and support with AI/ML-based capabilities, the following are the recommended segregation of future work -

- Identification of a series of patent search process phases that can be benefited from AI
- Assessment of AI/ML models which can help improve those phases.
- Articulation of AI research problems that need the resolution to perform those phases and activities
- Benchmarks, data sets, and other tools needed to support the research

5.2 Advancement Potential

Further deliberations and studies on this subject are of immense value at a global scale in the areas of

- Knowledge - development of knowledge and theory in the domains of intellectual property, innovation, and technology,
- Academia - development of academic subject and theory in the domains of intellectual property, artificial intelligence, and innovation diffusion,
- Practitioners - act as a guideline to evolve faster and alternative methods for IP Examination which can act as an accelerator in the process,
- Policy making - The findings of the proposed research shall be of immense use to policymakers and regulators to ensure transparent and reliable intellectual property protection practices,

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 1, October 2022
Impact Factor: 6.252

- Society - helps improve general awareness and understanding across private industry, government, and educational institutes.

REFERENCES

[1]. Lexology, March 2020, The benefits of Artificial Intelligence in the field of IP, Link Stephanie Crawford, Article - How to Do a Patent Search
[2]. Ayush Verma, July 2020, Artificial intelligence: Simplifying the process of prior art patent search, Link
[3]. Tyron Stading (Chief Data Officer at CPA Global), Blog - The Role of Artificial Intelligence in Intellectual Property, 2017
[4]. GreyB Services, 4 Types of Patent Searches: Everything You Need to Know
[5]. Leonidas Aristodemou, Et. Al., A literature review on artificial intelligence, machine learning and deep learning methods for analyzing intellectual property (IP) data, 2018
[6]. Brinda K. Varma, Treatise on Process Metallurgy: Industrial Processes, 2014, Link OECD, Patents, and Innovations: Trends and Policy Challenges, 2004, Link UpCounsel, Patent Search Techniques: Everything You Need to Know
[7]. Mahesh Maan, Article - Can We Together Solve the Major Challenges in AI Patent Search?, 2021, Link
[8]. David John, June 2015, Linked-in article – 'Major, but common Patent Searching Challenges', Link Poulomi Sen, July 2020, Artificial intelligence: Simplifying the process of prior art patent search, Link Benjamin Hanrahan, Esq., NA, How to Perform a Patent Search
[9]. TechFunnel Contributors, Nov 2019, 3 Emerging Technology Research & Development (R&D) Trends, Link
[10]. Ritesh Mathur, Linkedin Article - Challenges in Patent Analytics, 2019
[11]. Makoto Iwayama Et. Al. , Chapter 4 - Challenges in Patent Information Retrieval, 2021, Link Michael Cooney, Oct 2018, Gartner: Top 10 2019 tech trends you should know
[12]. Muhammad Imran, January 2022, Machine Learning Models Explained: Overview, Types & much more!
[13]. Sébastien Ragot et al., IP lifecycle management using blockchain and machine learning: Application to 3D printing data files, World Patent Information Vol 62