Journal of Intellectual Property Rights
Vol 30, July 2025, pp 499-504
DOI: 10.56042/jipr.v30i4.6906

Developing Metadata Standards for Efficient Searching of Indian Patent Information: A Plan

Abhishek Kumar[1†], Ashish Kumar Chauhan[1] and Dinesh K Gupta[2]

[1]Information and Library Network (INFLIBNET) Centre, Infocity, Gandhinagar — 382 007, India

[2]Department of Library and Information Science, Central University of Haryana, Mahendargarh — 123 031, India

Received: 25th November 2023; revised: 5th May 2025

The patent metadata standards for Indian academic institutions and research organizations have been investigated through the following objectives: identify patent metadata specific to Indian academia, conduct a comparative analysis of metadata structures in inPASS (Indian Patent Advanced Search System), PATENTSCOPE (WIPO) and Derwent Innovation, and propose a customized metadata scheme for Indian patent information services. The data were systematically extracted from these databases, primarily focusing on major fields like applicant names, classification codes, and administrative metadata. A quantitative comparison of field structures and user-centred accessibility was used. Initial analysis suggests the presence of pronounced structures in the metadata, especially concerning institutional affiliations and research domains within the Indian environment. It concludes with the proposed hybrid metadata schema spanning global standards and local needs to fill the gaps of discoverability for Indian academic users. It contributes to the current debate regardingthe accessibility of patent information in developing countries.

Keywords: Intellectual Property, Patents in India, Patent Metadata, Patent Database, inPASS

We are currently witnessing the knowledge economy where knowledge plays a role of critical resource for all production activities and a vital resource in all kinds of business organizations in the form of knowledge assets and converting valuable information protected legally in intellectual property.[1] As such, intellectual property (IP) is an intangible asset. The primary forms of IP are patents, copyrights, trademarks, and trade secrets. One of the quintessential measures of innovation output is patents, as patent indicators reflect the original intellectual property or performance of different countries, regions, technologies, institutes, organizations, firms, etc. The first Indian Patent was "An Efficient Punkah-Pulling Machine", invented by George Alfred DePenning in 1856. In India, Patent law was existed in 1911 when the Indian Patents and Designs Act, 1911 was enacted. After that, the Patents Act 1970 came into force in 1972, amending the existing law describing Patents in India.[2]

There remained unsatisfactory provision of electronic databases on Indian patent information; it is worth noting that Indian patent information can also be retrieved from the International Patent Documentation Center (INPADOC), which contains a collection of bibliographic data and legal status data of Indian patents.[3] A database is an organised collection of data, generally stored and accessed electronically through a computer system. It allows the facilities to quickly access, manage, modify, update, control, and organise data. Metadata is structured data information. Metadata gives relevant information about data and creates the path to reach the correct data, using the original primary data which is more valuable and meaningful. In 2015, a new electrotonic database called inPASS was launched. However, the information in the database is less suited to the changing information requirements of the researchers/scientists/faculty members/students working in Indian institutions of higher learning and research organisations.

The present study intends to find metadata for the new patent information database, especially for Indian educational institutions and organizations. Furthermore, it aims to develop a model of a patent information database by analysing metadata used in available standard databases for Indian educational institutions and organizations to fulfil the requirements of Indian academia and research community in a single space and search, it would successfully meet the needs of Indian academic patent information seekers following an international standard. The database will conform to the needs of Indian educational institutions and research organizations.

†Corresponding author: Email: abhishek@inflibnet.ac.in

Patent Search Services

There are many patent search databases at the international level; one such key database is IPO's PATENTSCOPE. It is a publicly available global patent search system. It provides access to information about new technologies, which are often disclosed for the first time in patent applications. The PATENTSCOPE database gives access to the international Patent Cooperation Treaty (PCT) applications in the full-text format on the day of publication and to patent documents of participating national and regional patent offices.

Derwent Innovation is an exclusive patent research application that provides access to globally trusted, curated patent intelligence and scientific literature. Information provider Clarivate Analytics creates the Derwent Innovation database which was previously part of Thomson Reuters.

Indian Patent Advanced Search System (inPASS), the patent search system started in 2015, provides full details of the patents, especially Indian Patents. The inPASS search interface shows some similarities to PATENTSCOPE. The similarities between PATENTSCOPE and inPASS can be found in the field combination search page, which is quite user-friendly compared to other patent search interfaces. inPASS allows a searcher to use several limitations, such as assignee name, applicant name, IPC code, application date, etc., to retrieve or filter search result sets.

Review of Literature

Bache (2011) argued that the effectiveness of access to patent information largely depends upon the information storage and retrieval system used by the patent information system. The patent search primarily ensures that everything relevant has been found. It often seeks to demonstrate that something does not exist and that there is a high cost of missing a relevant document; standard IR effectiveness measurement only reveals part of the truth.[4]

Richter and MacFarlane (2005) focused on the appropriateness of the metadata for retrieving user-centred results from the classification system. Considers that the use of metadata could significantly improve the classification of patents. However, the results vary on different automated classification mechanisms, namely, Gazette and PFA classifications.[3]

Clarke (2018) considered searching important for appropriate items and information from the database. He mentioned that patent searching is different from searching for other technical documents and that a dedicated mindset, strategy, and tactics need to be applied to the task, which is why the unique way patent databases are constructed. It also mentioned types of patent searches, keywords, titles and beyond the titles, classifications, names, patent families, and so on.[5]

Zavalina *et al.* (2017) suggested that metadata quality affects access to information resources, affecting the accuracy and memory of the discovery results. Metadata quality support is required for unique materials only available in a particularly specialized digital repository that provides access. An example of such unique and valuable materials is a digital historical patent. It also describes a case study of metadata management for the digital collection of American patents released in the 19th-Parambhik 20th century. The metadata for this collection discusses the results of the evaluation of metadata changes to support quality assurance. It presents plans for further research to understand metadata management and quality assurance and the role of metadata changes in these processes.[6]

Noh, Jo, and Lee (2015) have focused in this research on keyword approaches for implementing text-mining to patent data. The factors discussed include: elements of the patent documents to use for keyword selection, keyword selection methods to use, number of keywords to select, and how to convert the keyword selection results into an analysable data format. Furthermore, the research findings are expected to suggest helpful guidelines for selecting and processing keywords for patent analysis, further increasing the reliability and validity of research using text mining.[7]

Abood and Feltenberger (2018) focused on automated patent landscaping, a suggestion that mutually leverages human domain expertise, heuristics based on patent metadata, and machine learning to generate high-quality patent landscapes with minimal effort.[8]

Bogaard *et al.* (2019) discussed digital libraries, and it is helpful to explain how users search in a collection. Analysing search patterns can help them to improve the user interface, store management, and search algorithms. However, search patterns may differ widely in different parts of a collection. The main focus of this article is to describe how to identify these search patterns within a well-curated historical newspaper collection using the existing metadata.[1]

Whitman (2011) argued that the patent information seeker expects accurate information about the patent when searched from the available database. A free wiki-based resource called Intellogist is discussed in the paper. The wiki gives objective information and expert opinions about various tools containing patent information worldwide.[9]

Salampasis & Hanbury (2014) presented PerFedPat, an interactive patent search system based on the approach of federated search and the ezDL framework, mainly covering Espacenet, Google patents, Patentscope and the MAREC2 collection. It was also elaborated on the process of how this tool was developed. "Professional search in the patent domain usually needs both an analytical and an exploratory type of search, which is characterized more often, compared to fact-finding and question-answering web search, by recall-oriented information needs and sometimes by uncertainty and evolution or change of the information need."[10]

Singh, Chakraborty, and Vincent (2016) discussed central patent offices and their databases, including IPO, SIPO, JPO, EPO, KIPO, USPTO, and WIPO. It also examines their strength and weaknesses. Advocates that awareness of the search options available in different patent information databases needs to be understood by the seekers of the patent information to get qualitative results.[11]

Moholkar, Roy, and Buddle (2015) find that there are several databases which can be used to search for Indian patent information, including IPAIRS (Indian Patent Information Retrieval System), which is the official and free searching tool of the Indian Patent Office. The authors explore the completeness of the patent coverage and the accuracy of the information available in three subscription-based databases, namely CIPIS, INFULL, MCPaIRS and IPAIRS.[12]

Krestel *et al.* (2021) proposed the concept of deep learning for patent analysis due to the rising number of patents worldwide. It also describes the patent office's work for any one or more Patent classifications, Prior art search, Data analysis, patent examination and image analysis. Summarize the state-of-the-art techniques and describe how those are applied to various tasks in the patent domain.[13]

From the reviews discussed above it can be described that the Information Storage and Retrieval System plays a vital role in accessing to patent information and that should be user-centred rather than system centred. Needless to mention that patent searching is different from searching for other documents and users' approach to patent searching needs to be considered while designing a database. The metadata quality plays an important role in access to information regarding patents, affecting both precision and recall of search results. There are variations in regard to metadata used by different patent databases. As such, for making country specific patent information and retrieval there is a need to develop a schema that suits the requirement of a country like India and such need arises when the amount of intellectual works form the country is on rise.

Data Collection

In the process of data collection, in inPASS, the search fields of 'applicant's name', 'applicant's country' 'Jawaharlal Nehru University', and 'India' were used as keywords with following Boolean operator 'AND' in the field of 'published' and inPASS gave 26 results published patent details except one unspecific result (because of the first three words of the keyword).

The same was conducted in the field of 'granted,' and the result was 2 with zero non-specific results. In another database (PATENTSCOPE), in the search field of 'All name' and Office' 'Jawaharlal Nehru University' and 'India' was put as keywords with the assistance of Boolean operator (AND) and PATENTSCOPE gave 24 results along with two 'granted' results out of the total 24. Similarly, third database was Derwent Innovation, 'Jawaharlal Nehru University', '1990 to 2019' and 'India'(IN) as keywords were used in the search field of 'Assignee/Applicant', 'Publication year' and 'Country code' bearing assistance of Boolean operators (AND) and as per the result 26 'published' and 5 'granted' were there out of 31 patent data. The detailed result sheet is given below in Table1.

Based on collected metadata results from PATENTSCOPE, Derwent Innovation, and inPASS, metadata were collected and were taken for evaluation in the Table 2. For the present research, only a few efficient metadata fields were selected from the Derwent Innovation, in which more than 197 fields are available, out of which the most useful few have been chosen. Except for a few explicit differences, all of them are quite similar. Differences include in the nomenclature quantity of information. To be specific, categories are differently entitled. For example, in the case of inPASS, the entitlements are pretty specific but

Table 1 — Brief of data extracted from three different databases

S. No.	Key Word	Database	Published Patent(PP)/Granted Patent(PG)	Results
1	1:ApplicantName 2: Applicant Country "Jawaharlal Nehru University" (AND) India	In PASS	PP PG	26 2
2	1.All Name 2.Office "Jawaharlal Nehru University" AND India	PATENTSCOPE	PP 24 / PG 2	24
3	1.Assignee/Applicant 2.Publication Year 3.Country Code “Jawaharlal Nehru University” AND 1900 to 2019 AND India (IN)	Derwent Innovation	PP 26 /PG 5	31

Table 2 — Patent metadata from three different databases

A	B	C	D	E
Derwent Innovation	PATENTSCOPE	inPASS_ Publish	inPASS_ Granted	Auxiliary Fields
Publication Number	Office	Invention Title	Grant Title	AISHE code
Title	Title	Publication Number	Patent Number	Source
Title– DWPI	Application Number	Publication Date	Application Number	Name Variations
Priority Number	Application Date	Publication Type	Typeof Application	Date of Renewed
Priority Date	Publication Number	Application Number	Parent Application Number	Allied Subject/Field
Priority Date – DWPI	Publication Date	ApplicationFiling Date	Date ofPatent	
Application Number	Publication Kind	PriorityNumber	Date ofGrant	
Application Date	IPC	PriorityCountry	Date ofRecordal	
Publication Date	Applicants	PriorityDate	AppropriateOffice	
EstimatedExpiration Date	Inventors	Field OfInvention		
IPC – Current	Priority Data	Classification (IPC)		
Assignee/Applicant	Grant Number	InventorName		
OptimizedAssignee	Grant Date	Address		
IPC Class	Country	Country		
IPC ClassGroup	Nationality	Nationality		
IPC Subclass	Applicant Name	Applicant Name		
IPC Subgroup	Address	Address		
IPC Current Full	Country	Country		
Language of Publication	Nationality	Nationality		

Table 3 ─ Metadata fields for the patent retrieval

FIELD(S)	METADATA FIELD	A	B	C	D	E	DESCRIPTION
A+B+C	Patent Title	✓	✓	✓	-	-	*Patents, which protect ownership of invention or discovery of a new and valuable process like the machine or mechanism, or composition of matter or any new and useful development of those items.*
A+B+C	Application No.	✓	✓	✓	-	-	*The application number for the patent.*
A+B+C	Date of application	✓	✓	✓	-	-	*The application date of the patent.*
A+B+C	Publication No./Journal No.	✓	✓	✓	-	-	*On its publication on patent journal, the patent receives a journal no/publication no.*
A+B+C	Published Date	✓	✓	✓	-	-	*Date of publication*
D	Grant Title	-	-	-	✓	-	*When the patent is granted, it confirms a title as the final.*
B+D	Patent Grant No.	-	✓	-	✓	-	*The complete identifying number attached to the patent.*
B+D	Granted Date	-	✓	-	✓	-	*Date of being granted.*
B+C	Inventor Name (s)	-	✓	✓	-	-	*Name of the inventor*
C	Address	-	-	✓	-	-	*Address of Inventor*
C	Country	-	-	✓	-	-	*Country of residence of the inventor(s)*
C	Nationality	-	-	✓	-	-	*Inventor's nationality*
A+B+C	Applicant Name (s)	✓	✓	✓	-	-	*Name of the individual/organisation/institution applied for*
E	AISHE Code	-	-	-	-	✓	*MHRD (All India Survey of Higher Education) is allotted a code known as AISHE code*
E	Name Variation	-	-	-	-	✓	*Name of institute with all variation used and old name*
A+B+C	Address	✓	✓	✓	-	-	*Address of applicants*
B+C	Country	-	✓	✓	-	-	*Country of residence of the applicant(s)*
B+C	Nationality	-	✓	✓	-	-	*Applicant's nationality*
B+C	Subject Field	-	✓	✓	-	-	*Subject of patent*
A+B+C	IPC No.	✓	✓	✓	-	-	*International Patent Classification (IPC)*
B	Office of filling	-	✓	-	-	-	*National and Regional application authority's names*
E	Date of Renewed	-	-	-	-	✓	*Date of renewal of the patent*
A+B+C	Language	✓	✓	✓	-	-	*The language of the original item*
E	Source	-	-	-	-	✓	*From where the patent data has been taken*
E	Allied Subject/Field	-	-	-	-	✓	*Other relevant subject/filed where the Patent is applicable*

need to be more user- friendly (especially for new users or non-experts) in comparison PATENTSCOPE. If compared with the former two patent databases, Derwent Innovation is a paid service offering many unique outputs like specific subject categories, titles, and classification numbers, etc. However, the present study has prioritised efficient fields for a better outcome and to build a more efficient database structure for the researchers of country like India.

Data Analysis and Results

In Table 2, column number "A" represents the metadata taken from Derwent Innovation; similarly, column "B" represents PATENTSCOPE. Besides columns A and B, columns "C" represent inPASS metadata for published patents, and "D" represents inPASS metadata for granted patents. The last column, "E", contains a few auxiliary relevant metadata from this study.

During the search in these databases, few problems or inconveniences were encountered while finding the outcomes. For example, the option to download data is only available in Derwent Innovations and not inPASS or PATENTSCOPE. Consequently, it is time-consuming and violates the accuracy of data retrieval. In inPASS, many specific inputs are required to retrieve maximum data, which is unlikely in Derwent Innovation and PATENTSCOPE. While inPASS and PATENTSCOPE are open sources for data retrieval, Derwent Innovation is a highly paid service, and only some organizations can afford it despite being immensely suitable. However useful these three databases are, the most efficient fields/categories have been chosen for the present requirement to configure a better data retrieval system particularly suitable for offering Indian patent information services to the patent information seekers.

Findings

As part of the findings, the Table 3 is prepared according to the relevance and convenience based on evaluating the collected metadata from the above

sources (Table 2). The Table 3 contains an analytical illustration of the metadata sources. The representation in Table 2 broadly resonates with the understanding of the table. The category of 'field' (in Table 3) represents the homogeneity in their operation in combined form or individually. Each category is placed according to its efficiency and user-friendly potential. Few additions – (1) AISHE code, which represents the MHRD rewarded code of academic institutions to retrieve specific data about the institution; (2) Source, which represents the exact information about the source of the patent with authentic data; (3) Name Variation, it covered all names of the institute with the old name and (4) Allied Subject/Field.

Conclusion

From the review, this study suggests that many researchers are concerned about various aspects of patent information, resulting in supporting quality searches for final users. The existing patent database needs to meet the requirements of each user community. In the digital age, many participants require more help. However, inappropriate keyword reactions may reduce users' confidence despite abundant data. To address this, an important step is to design a user-friendly database that can effectively respond to the keyword by evaluating the metadata and user's tendency. Building a database by collecting metadata (as shown above) will be helpful for a few practical reasons. Beginning with academic institutions and research organisations, researchers and learners will find it quite friendly and convenient. Of course, a few valuable databases like inPASS, PATENTSCOPE, Derwent Innovation, Web of Science, Scopus, and PubMed provide detailed information. However, these do not fully support the convenience or expectations of the Indian patent information seeker.

Therefore, an attempt has been made to design a metadata structure for the patent database for researchers in a country like India. As per the objective, a thorough analysis of three patent databases (inPASS, Derwent Innovation, and PATENTSCOPE) was undertaken with substantial fruitful insights to develop a model for Indian information seekers. Primarily, the present model deals with bibliographical metadata. This model aims to be suitable for Indian patent information seekers so that an organization or institution can develop its patent database. On successful installation of this model, it has a future scope of evolving from a bibliographical database to a full-text patent database (containing content and all). This is only the blueprint of the structure, which requires substantial resources to grow and develop. This proposed database is patterned to suit the needs of Indian patent information users, and Indian researchers will retrieve the information.

Reference

1 Bogaard T, Hollink L, Wielemaker J, van Ossenbruggen J & Hardman L, Metadata categorization for identifying search patterns in a digital library*, Journal of Documentation,*75(2) (2019) 270, https://doi.org/10.1108/JD-06-2018-0087.

2 Vaish Associates Advocates, (2019), India: Patents Law In India – Everything You Must Know, From Mondaq: connecting knowledge &people:http://www.mondaq.com/india/x/125766/Patent/Patents+Law+In+India+Everything+You+Must+Know

3 Richter G & MacFarlane A, The impact of metadata on the accuracy of automated patent classification, *World Patent Information*, 27 (1) (2005) 13.

4 Bache R, Patent retrieval–A question of access, *World Patent Information*, 33 (4) (2011) 345.

5 Clarke N S, The basics of patent searching, *World Patent Information*, 54, S4-S10. Databases, *World Patent Information*, 41 (2018) 38.

6 Zavalina O L, Phillips M & Tarver H S, Quality assurance and evaluation of change for patent metadata: Quality Assurance and Evaluation of Change for Patent Metadata, *Proceedings of the Association for Information Science and Technology*, 54 (1) (2017) 842, https://doi.org/10.1002/pra2.2017.14505401180.

7 Noh H, Jo Y & Lee S, Keyword selection and processing strategy for applying text mining to patent analysis, *Expert Systems with Applications*, 42 (9) (2015) 4348, https://doi.org/10.1016/j.eswa.2015.01.050.

8 Abood A & Feltenberger D, Automated patent landscaping, *Artificial Intelligence and Law,* 26(2) (2018) 103, https://doi.org/10.1007/s10506-018-9222-4.

9 Whitman K, Intellogist: an online community dedicated to comparing major patent search systems, *World Patent Information*, 33(2) (2011) 168.

10 Salampasis M & Hanbury A, PerFed Pat: An integrated federated system for patent search, *World Patent Information*, 38 (2014) 4.

11 Singh V, Chakraborty K & Lavina-Vincent C, Patent Database: Their Importance In Prior Art Documentation and Patent Search, *Journal of Intellectual Property Rights*, 21 (2016) 2.

12 Moholkar M, Roy S & Buddle K, Comparative study of Indian patent databases, *World Patent Information*, 41 (2015) 38, https://doi.org/10.1016/j.wpi.2015.03.001.

13 Krestel R, Chikkamath R, Hewel C & Risch J, A survey on deep learning for patent analysis, *World Patent Information*, 65 (2021) 102035.