Task:

1. Deep research on how to get bulk data from Indian patents. Is there a way to download bulk data without webscrapping?

Reference: <https://iprsearch.ipindia.gov.in/publicsearch> , <https://ipindia.gov.in/patents.htm>

Update 1: 10th june 2025

1. Public Search Portal (webscrapping)

<https://iprsearch.ipindia.gov.in/publicsearch>

Allows search by application/publication number, applicant/inventor name, keywords, IPC, dates.

Downloadable PDFs: specifications, application details, FERs, etc.

Requires Selenium for automation due to dynamic form elements.

1. IP India Patents Portal

<https://ipindia.gov.in/patents.htm>

Provides procedural info, forms, legal notices, and links to filing systems.

How to file patent and stuff

Not really helpful,

1. Weekly Patent Journals (doc scrapping)

<https://search.ipindia.gov.in/IPOJournal/Journal/Patent>

Abstracts, application numbers, statuses published weekly as PDFs.

Can be downloaded and parsed using tools like pdfplumber, PyMuPDF, or OCR.

1. WIPO PATENTSCOPE Access

<https://www.wipo.int/en/web/patentscope/w/news/2018/news_0002>

Indian patent data is available via WIPO’s global patent search system.

XML-format bulk data access possible through WIPO’s data services.

WIPO PATENTSCOPE: https://patentscope.wipo.int/search/en/search.jsf

1. Archived StackOverflow Reference on Web API

Historical info on undocumented Indian patent API (now disabled):

<https://stackoverflow.com/questions/39356677/webscraping-the-indian-patent-website-for-patent-data>

1. Institutional Access / RTI Route

Data access through academic partnerships or RTI requests is possible.

Some vendors may offer structured datasets under formal agreements.

Note: The public search robots.txt is currently not accessible, and the domain www.search.ipindia.gov.in returns certificate errors.

Update 2: 11th june 2025

1. Script that automates the scraping of patent data from the Indian IP Office's public search portal: <https://iprsearch.ipindia.gov.in/publicsearch>

On a monthly basis (defined by user) and saves it in a csv

Note: need to manually input the captcha

1. Script to download the latest patent journal and read the data from the pdf.

1. Saturo Global’s Indian Patent Database

Provides PDF records from 1912 and curated XML data from 2013 onward

Updated every 15 days with latest filings

Offers bulk downloads via APIs or structured [fileshttps://saturoglobal.com/products/indian-patent-database/](http://saturoglobal.com/products/indian-patent-database/)

Update 3: 13th june 2025

1. Script that automates downloading of pdfs (complete specification, forms, legal documents, etc) from the <https://iprsearch.ipindia.gov.in/publicsearch> website for desired dates.

* Downloading pdfs for 5 applications took around 3 minutes

Update 4: 16th june 2025

1. Contents of the PDF’s (applied):

* Complete Specifications: FIELD OF THE INVENTION, Background of the Invention, How This Invention Addresses the Gaps, Detailed description of the invention with accompanying drawings, Industrial Applicability, Claims, Abstract
* Declaration as to Inventorship: Applicants, Declaration, Inventor(s)
* Form 1: Detailed information of applicants and inventors (Name Nationality Address Country State District City), . TITLE OF THE INVENTION, ADDRESS FOR CORRESPONDENCE OF APPLICANT / AUTHORISED PATENT AGENT IN INDIA, PARTICULARS FOR FILING PATENT

1. Contents of the PDF’s (granted):

* Intimation of Grant: application no., Address of Service, Email Id, Patent No, Name Of Applicant, Date of Patent, Priority Date, Filling date of Request for examination, Title, Number of claims, Controller' Name, Controller' Email
* Patent Certificate: Patent No, Application No, Date of Filing, Patentee, Granted Date
* Form 3: All the countries that company has applied for/been granted a patent , PCT

Update 5: 24th june 2025

* Script to automate downloading of the required documents and extracting Abstract, Applicant(s), Application No, Field of Search, Filing Date, Inventor(s), Patent Number, Priority Date, Publication Date and Title from the documents.

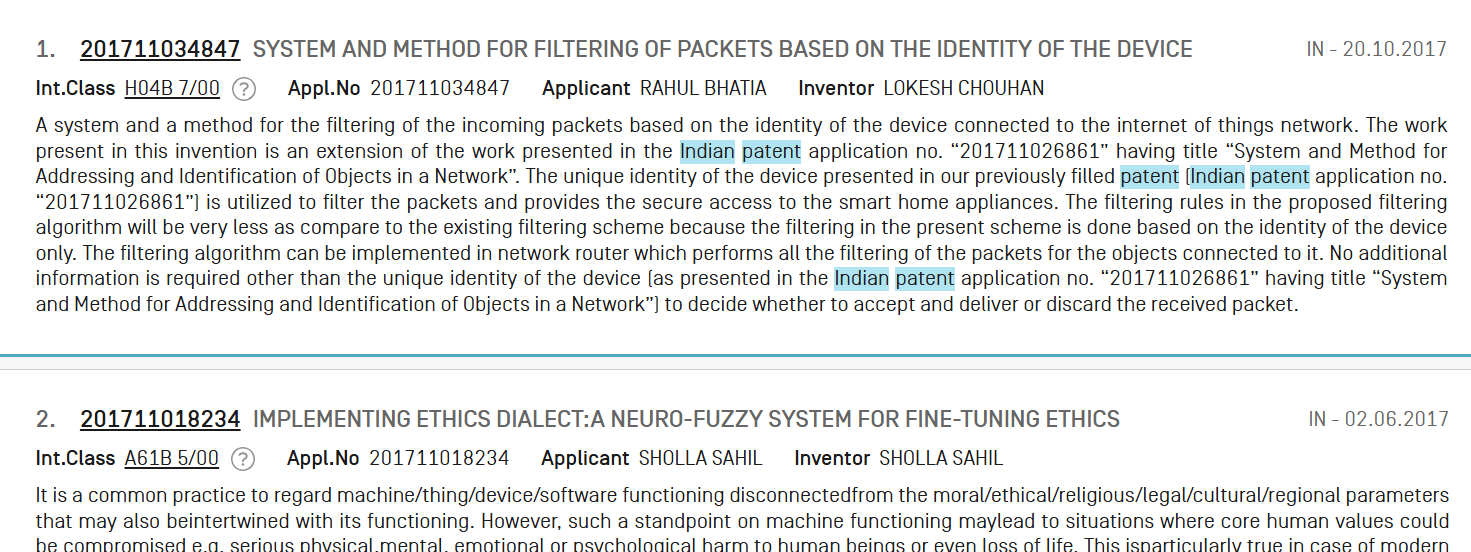
Update 6: 26th june 2025

* Check if patent no of indian patent related to wipo , If there is any relation between ipo and wipo, Is application no the doc\_id in wipo, if there is some relation with wipo

| **Connection Type** | **Description** | **Example Format** | **Where Found** |
| --- | --- | --- | --- |
| **Priority Number** | A WIPO application claims priority from an Indian filing | IN202341012345 | WIPO → Priority Data section |
| **International Application Number** | An Indian application is a **national phase entry** from a WIPO PCT application | PCT/IN2023/012345 | IPO complete spec / Form 1 |
| **WO Publication Number** | Used to identify international (WIPO) patent application | WO2023/123456 | WIPO → doc\_number |
| **National Phase Entry** | Indian application entered from a WIPO PCT filing | 2023112345 (IPO format) | WIPO → Designated States → IN |
| **Applicant/Inventor Matching** | Names matched across IPO and WIPO systems | Dr. ABC, XYZ Institute | Metadata fields: applicant, inventor |
| **Title & Abstract** | Often matched between IPO and WIPO for cross-referencing | Title/abstract text | WIPO + IPO publications |
| **Publication Dates** | IPO publishes after national phase entry from WIPO | DD-MM-YYYY | Both IPO & WIPO records |
| **IPC Classifications** | Shared international classification codes | A61K 8/00, G06F 17/30 | Common to both IPO and WIPO metadata |
| **PCT Filing via IPO** | IPO acts as a Receiving Office for Indian-originated PCT applications | PCT filed through IPO | Submission acknowledgment + WIPO listing |

You can search for the patent using application number itself from ipo website:

<https://patentscope.wipo.int/search/en/result.jsf?_vid=P20-MCEX73-45941>



Update 7: August 2025

Github link : <https://github.com/khushi-inventohub/Indianpatentoffice>

Contains the code to scrape data from the <https://iprsearch.ipindia.gov.in/publicsearch> website and extract the information and save the extracted files and extracted data in json and paraquet file in aws S3 bucket (month wise).