

Automation of Patent Searches at INPI: A Web Scraping and Data Analysis Approach

João Alexandre Brito^{1*}, Celso Barreto da Silva², Fabio Fonseca Barbosa Gomes²

¹Startup Innovation Hub of SENAI CIMATEC University; ²SENAI CIMATEC University, Postgraduation Department; Salvador, Bahia – Brazil

Patent searches are an essential step in technological development and innovation, but the traditional search process can be time-consuming and inefficient. This article presents PATENTIA, a system that automates patent searches at the National Institute of Industrial Property (INPI) using web scraping techniques and relational database data storage. The system allows the separation and analysis of patent data, presenting it in an organized way in a web interface developed with React and Node.js. In addition, the tool allows the export of results in Excel format, making it a useful resource for researchers, inventors, and companies. This study details the development of the tool, the challenges encountered, and the benefits of automation in the patent search process. **Keywords:** Patent Search. Search Automation. Web Scraping. INPI. WIPO.

Patent research is a fundamental process for researchers, inventors, and companies seeking to assess the originality of an invention and ensure its legal protection. According to Souza and Almeida (2020) [1], access to information contained in patent databases is essential for the development of technological innovation. Similarly, Santos (2018) [2] emphasizes that the use of patent databases allows for strategic market and competitor monitoring. In Brazil, the main database for patent searches is the National Institute of Industrial Property (INPI), which provides information on patent applications and grants.

The traditional search process on the INPI [3] platform can be time-consuming and inefficient, requiring users to perform repetitive manual searches to obtain relevant information. As highlighted by Lima and colleagues (2019) [4], the large volume of data and the lack of advanced search tools hinder the analysis and extraction of relevant patent information. Oliveira and Mendes (2020) [5] point out that automating data collection and processing can significantly accelerate the retrieval of essential information.

To address this issue, the present study proposes the automation of patent research through the tool PATENTIA. This solution is based on the use of web scraping—a technique for extracting data from the web—allowing for the automatic collection of information from INPI [3]. According to Silva and Ferreira (2021) [6], web scraping has been widely used to automate repetitive tasks involving large volumes of data, making previously slow processes more efficient. Likewise, Costa and Ribeiro (2022) [7] argue that the application of scraping techniques can drastically reduce the time spent collecting information from public databases.

PATENTIA aims to implement an intuitive web interface for patent searches, enabling users to perform queries, view organized information, and export results in Excel format for further analysis. According to Almeida (2021) [8], intuitive interfaces significantly enhance user efficiency when interacting with large data sets. Pereira (2020) [9] also emphasizes that visualization and data filtering tools allow for more objective and structured analyses, contributing to more assertive decision-making.

PATENTIA uses the Python programming language for automating the data collection process, Node.js for backend management, and the React language for the frontend, ensuring interactivity and usability of the system. According to Machado and Lopes (2022) [10], the combined

Received on 12 February 2026; revised 18 April 2026.

Address for correspondence: João Alexandre Brito. SENAI CIMATEC University. Av. Orlando Gomes, 1845 - Piatã. Salvador, Bahia, Brazil Zipcode: 40650-010. E-mail: joao.brito@fieb.org.br.

J Bioeng. Tech. Health 2026;9(5):383-391
© 2026 by SENAI CIMATEC University. All rights reserved.

use of modern technologies such as Python and Node.js has enabled the creation of faster and more scalable solutions for information extraction and processing. Supporting this idea, Nogueira (2021) [11] argues that choosing frameworks like React provides more responsive and intuitive interfaces, enhancing the end-user experience.

In this context, the objective of this article is to present the development of the PATENTIA search tool, discuss the challenges of automating patent searches, and assess the impact of the tool on access to technological information available through INPI. The research is structured in sections detailing the methods used, the implementation of the tool, and the results obtained. As stated by Santos and Carvalho (2022) [12], the implementation of digital tools in patent knowledge management contributes to the optimization of industrial and academic processes. Thus, the PATENTIA solution represents an important step toward modernizing patent research in Brazil, as also emphasized by Vieira and Rocha (2023) [13], who highlight the positive impact of digitizing bureaucratic processes on research and technological innovation.

Theoretical Framework

Robotic Process Automation (RPA) emerges as an innovative solution for optimizing routine tasks, reducing human error, and increasing operational efficiency. According to Silva and colleagues (2023) [14], RPA enables companies to automate their most repetitive administrative processes, such as bill payments and report generation, freeing human resources for more strategic activities.

Furthermore, studies such as Freitas (2023) [15] highlight RPA's ability to integrate data from multiple systems, promoting a consolidated and reliable view of organizational processes. Thus, the application of RPA transcends sectors, being implemented in both corporate environments and governmental institutions, as emphasized by Souza and Almeida (2020) [1].

The databases of the National Institute of Industrial Property (INPI) and the World Intellectual Property Organization (WIPO) are fundamental sources of information for intellectual property analysis; however, they present challenges related to data volume and access complexity. The study by Silva and colleagues (2023) [14] points out that the INPI database is designed to support public queries but lacks advanced integration and visualization tools. Additionally, according to Santos and Carvalho (2023) [12], WIPO provides comprehensive patent classification and description resources, although the absence of automated processes still limits query efficiency. Integrating these databases with RPA technology, as advocated by Lima (2023) [16], can transform the user experience by enabling faster and more accurate access.

Data quality and reliability are critical factors for decision-making, especially in processes involving intellectual property. Silva and colleagues (2022) [14] emphasize that RPA application can mitigate errors caused by human intervention, promoting a more trustworthy data environment. Moreover, studies conducted by Costa (2023) [17] demonstrate that automation contributes to the validation and standardization of captured information, ensuring greater consistency. Finally, Oliveira (2023) [18] reinforces that the use of automated analysis tools significantly improves the accuracy and speed of data interpretation.

Data visualization is an essential element for understanding complex information, especially in scenarios involving large volumes of records. According to Freitas (2023) [15], interactive dashboards offer an efficient approach to presenting data, helping to identify patterns and relationships. Likewise, Santos and Carvalho (2023) [12] highlight that implementing graphical tools in intellectual property databases enables faster and more intuitive interpretation of results. Additionally, Lima (2023) [16] suggests that the use of visual elements, such as interactive maps, can enhance communication between different areas of an organization.

Data integration and time optimization are critical challenges that can be addressed through process automation. Costa (2023) [17] points out that using RPA for dynamic data querying provides a unified and centralized view, reducing information dispersion. Complementarily, the study by Oliveira (2023) [18] indicates that this centralization can reduce up to 40% of the time spent on manual queries. According to Santos e Carvalho (2023) [12], this agility enables greater focus on high-value activities, such as the strategic analysis of collected data.

Practical applications of RPA for intellectual property data analysis also include the possibility of creating search histories and pagination functionality, as demonstrated by Silva and colleagues (2023) [14]. These tools facilitate navigation through large volumes of records, allowing for quick retrieval of previously accessed information. Additionally, Santos e Carvalho (2023) [12] observe that search history promotes the reuse of data in future analyses, increasing productivity. Freitas (2023) [15] adds that proper pagination ensures a more efficient experience for the end user, eliminating performance bottlenecks during navigation.

These aspects, when integrated, provide a robust theoretical foundation for developing RPA-based solutions capable of significantly transforming how intellectual property databases are queried and analyzed. The combination of automation, efficient visualization, and data integration represents a significant advancement toward optimizing processes related to intellectual property.

Materials and Methods

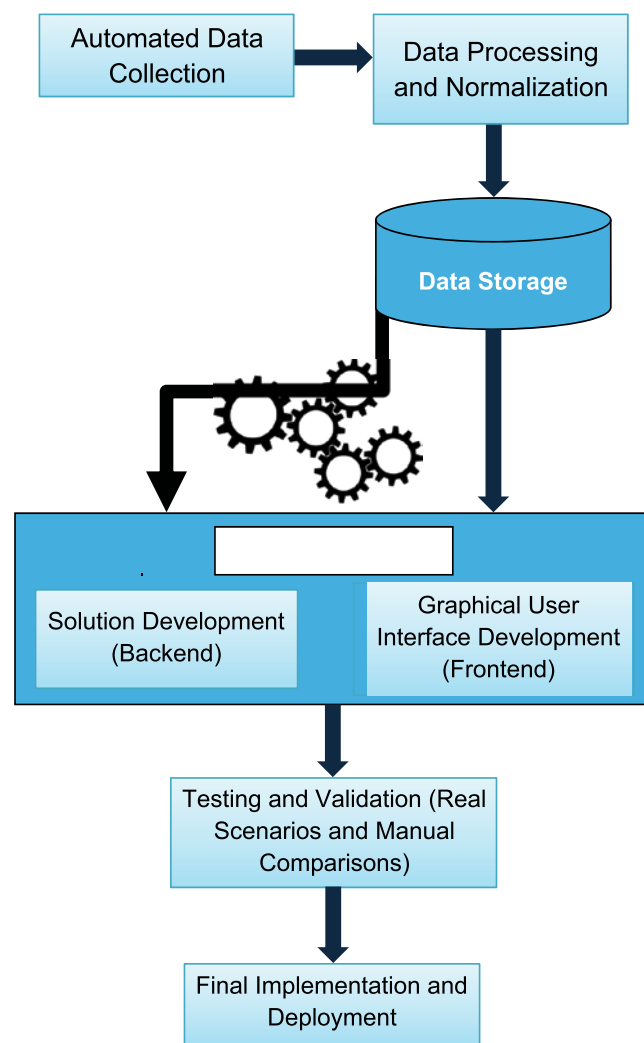
The methodology employed in this study followed a structured scientific approach, encompassing essential steps to ensure the rigor, validity, and replicability of the results obtained during the development of the PATENTIA system. Initially, a systematic literature review was conducted to identify and understand the existing challenges in manual patent searches, providing

a theoretical basis for the need for automation, as highlighted by authors such as Lima and colleagues (2019) [4] and Oliveira and Mendes (2020) [5].

The adopted scientific methodology included the clear formulation of the problem, definition of specific objectives, appropriate selection of technologies and tools, data collection and processing, implementation of the solution, execution of empirical tests and validations, as well as analysis and discussion of the results, as illustrated in Figure 1.

The development of the system was structured into the following methodological steps, as detailed in Figure 1:

Figure 1. Methodological diagram.



- 1. Literature Review and Problem Definition:** Identification of issues and challenges related to manual patent searches through critical analysis of existing literature.
- 2. Selection of Technologies and Tools:** Informed selection of modern and efficient technologies such as Python, Node.js, React, MySQL, Selenium, and Requests to meet the specific needs of the project.
- 3. Automated Data Collection (Web Scraping):** Implementation of the Web Scraping technique using Python, Selenium, and Requests, ensuring robustness and efficiency in extracting data directly from the INPI database.
- 4. Data Processing and Normalization:** Use of Python scripts to ensure the standardization, consistency, and quality of the collected data, eliminating duplicates and errors.
- 5. Data Storage and Management:** Storage of processed data in a MySQL database, using Sequelize as the ORM framework to ensure performance and scalability.
- 6. Backend and Frontend Development:** Construction of a robust and scalable platform with a backend developed in Node.js and an interactive frontend built with React, offering a user-friendly and functional interface.
- 7. Testing and Validation:** Evaluation of the tool's accuracy and efficiency through empirical testing and systematic comparisons between manually and automatically obtained results.

What Problem Does This Project Solve

PATENTIA aims to address the limitations associated with manual patent searches within the database of the National Institute of Industrial Property (INPI). Currently, researchers and companies must manually access this database, perform repetitive queries, and extract information in an unstructured manner. According to INPI (2024) [3], conducting a patent search requires users to be familiar with classification codes and perform refined queries to obtain relevant results. Furthermore, Alvarez, Filgueiras, and Angeli

(2023) [19] point out that identifying public domain patents in Brazil poses challenges due to the data structure and the need for repeated searches to locate specific information.

The PATENTIA system automates this process, enabling fast and accurate searches while organizing data in a user-friendly interface with export options to spreadsheets, making patent access and analysis significantly more effective.

According to Silva, Santos, and Silva (2021) [14], platforms such as Patentscope (WIPO), Espacenet (EPO), and BuscaWeb (INPI/BR) already offer free patent search tools, but the structure in which information is presented can still be improved to facilitate user analysis.

Comparative Analysis of Patent Databases

Patent searching is a crucial step in the process of innovation and technological development. Several databases are available to support researchers and companies in this task, each with specific features that can influence the efficiency and accuracy of the searches. Among the main patent databases are: Espacenet, Patentscope, Google Patents, Lens, Derwent Innovation Index, and Orbit Intelligence.

One of the main difficulties faced by users of these databases is the complexity involved in locating and interpreting the desired options and patents, as demonstrated in Figures 2 and 3 from the INPI database. This is partly due to the need for familiarity with specific classification codes and the requirement to conduct refined searches to obtain relevant results. Furthermore, patent descriptions may be ambiguous or generic, making the identification process more challenging. Alvarez and colleagues (2023) [19] emphasize that identifying public domain patents in Brazil faces difficulties due to the data structure and the need for repeated queries to find specific information.

Figure 2 presents the initial interface of the database provided by the National Institute of Industrial Property (INPI), which offers access to various categories related to industrial property,

such as trademarks, patents, industrial designs, geographical indications, computer programs, integrated circuit topographies, technology transfer, and patent-related technological information.

However, it is observed that this interface does not provide support elements or clear instructions to facilitate navigation and understanding for beginner or inexperienced users in intellectual

property searches. This lack of guidance may hinder the research process, reinforcing the need for more intuitive and interactive alternative tools, such as the PATENTIA system.

Therefore, the standard INPI interface highlights the importance of solutions that enhance usability and accessibility, especially for audiences less familiar with the process of searching for and analyzing patents.

Figure 2. INPI database interface.



Source: Instituto Nacional de Propriedade Industrial (INPI) [3].

Figure 3. INPI database.



Source: Instituto Nacional de Propriedade Industrial (INPI) [3].

Each of the aforementioned databases has distinct features that can meet different user needs. For example, Espacenet, developed by the European Patent Office (EPO), offers free access to more than 90 million patent documents and is recognized for its user-friendly interface and comprehensive coverage (Figures 4, 5 and 6). Patentscope, maintained by the World Intellectual Property Organization (WIPO), also provides free access and includes advanced

search functionalities, such as the ability to search for chemical sequences.

Google Patents integrates information from various databases and features a simplified interface, making it more accessible for less experienced users. Lens stands out by combining patent data with academic information, making it especially useful for research involving scientific innovation. On the other hand, Derwent Innovation

Figure 4. Interface of the ESPACENET database.



Source: ESPACENET (2025).

Figure 5. Classification-based search in the ESPACENET database.



Figure 6. IPC code search in the ESPACENET database

Source: ESPACENET (2025).

Index and Orbit Intelligence are paid platforms that offer advanced data analysis and visualization tools. These platforms are widely used in corporate and academic environments for competitor monitoring and identifying technological partners.

Table 1 presents a comparison of the main features of these databases. The choice of the most suitable database will depend on the specific needs of the user, considering factors such as the depth of the research, available resources, and familiarity with the tools offered by each platform.

Results and Discussion

The implementation of the PATENTIA system enabled significant improvements in the process of patent search and data analysis. Among the main results observed, several stand out due to their impact on efficiency, accuracy, and usability. One of the key achievements was the standardization of the search process. The system ensures that every patent search follows the same structured

steps, eliminating inconsistencies that may arise from manual searches. This uniformity contributes to greater reliability and repeatability of results, especially when dealing with large volumes of information.

Moreover, PATENTIA allowed an increase in the volume of data retrieved per search within a reduced time frame. Traditional patent queries often require manual access and filtering through multiple pages and parameters. With the automated web scraping and structured database integration, the system is capable of processing dozens of patent records in a fraction of the time previously required, significantly boosting productivity.

Another important result was the development of a timeline-based visualization for the retrieved data. Through a dynamic dashboard interface, users can visualize patent filings and trends across different time intervals. This feature enhances the strategic analysis of intellectual property landscapes, providing intuitive insights that support decision-making processes. The tool also features

Table 1. Comparison of the main patent database.

Database	Access	Coverage	Search Tools	Statistical Analysis
Espacenet	Free	Over 90 million patents	User-friendly interface, keyword search, and extensive classification; includes patent family analysis and global coverage	Limited
Patentscope	Free	International and national patents	Advanced search, including chemical sequences	Limited
Google Patents	Free	Integrates multiple databases	Simplified interface, keyword search	Limited
Lens	Free	Patents and academic literature	Keyword and classification search, integration with academic data	Moderate
Derwent Innovation Index	Paid	Extensive, with patent family analysis	Advanced search and analysis tools	Advanced
Orbit Intelligence	Paid	Comprehensive, with global coverage	Advanced tools for search, analysis, and data visualization	Advanced

an Excel export function, allowing users to generate spreadsheets with structured patent data. This makes it easier to conduct external analysis, share information with third parties, and incorporate the results into business or academic reports.

Furthermore, the system proved to be effective in reducing human error during the search and analysis stages. Manual searches are prone to data omissions, incorrect filtering, and transcription mistakes. By automating the entire pipeline—from data collection to presentation—PATENTIA minimizes these risks and ensures greater data integrity.

In addition to these main points, the architecture of PATENTIA, built using modern technologies such as Python, Node.js, and React, enabled the creation of a responsive and user-friendly interface, which proved to be a crucial element in improving the overall user experience.

Conclusion

This study presented the development and implementation of PATENTIA, a system designed

to automate patent searches at the National Institute of Industrial Property (INPI) using web scraping techniques and modern web technologies. The solution addresses several inefficiencies associated with the traditional manual search process, offering a reliable, scalable, and user-friendly alternative for researchers, inventors, and organizations engaged in innovation.

The automation of the search process through standardized procedures ensures consistency and reduces dependency on user expertise, while significantly increasing the volume of data processed in less time. This improvement has a direct impact on productivity and strategic intelligence, particularly in R&D environments where timely access to information is essential for innovation and competitive advantage.

One of PATENTIA's most notable contributions is the transformation of complex and disorganized datasets into structured, easily interpretable information. The integration of a timeline-based visualization dashboard enables users to identify historical trends, monitor technological development, and anticipate market movements based on patent

activity. Additionally, the Excel export function enhances usability by enabling further analysis and data sharing across different platforms.

Another major impact of the system is the reduction of human error in data retrieval and processing. By automating repetitive and error-prone tasks, PATENTIA enhances the reliability of patent data and allows professionals to focus on higher-value activities, such as strategic planning and innovation design.

Furthermore, the project highlights the relevance of combining emerging technologies—such as Python for automation, React for intuitive user interfaces, and Node.js for backend processing—to create effective solutions for public data access. The modular and extensible architecture of PATENTIA also opens possibilities for future improvements, including integration with other patent databases (such as WIPO and EPO), implementation of AI-driven search optimization, and multilingual support.

In conclusion, PATENTIA represents a significant advancement in the way intellectual property data is accessed and used in Brazil. It offers not only technical innovation but also a strategic tool that democratizes access to patent information, promotes efficiency, and supports the broader goals of innovation and technological development. The solution has the potential to inspire further initiatives in digital transformation within the intellectual property domain, setting a new standard for transparency, accessibility, and intelligence in patent research.

References

1. Souza A, Almeida B. O acesso à informação em bases de patentes e seu impacto na inovação tecnológica. *Rev Bras Inov.* 2020;2:45-60.
2. Santos F. Monitoramento estratégico de mercado por meio da análise de patentes. *Rev Est Propr Intelect.* 2018;1:30-40.
3. Instituto Nacional da Propriedade Industrial (INPI). Guia prático para buscas de patentes. Rio de Janeiro: INPI; 2025. Available from: <https://www.gov.br/inpi>
4. Lima C, et al. Dificuldades de acesso a bases de dados de patentes: um estudo sobre o INPI. *Cienc Tecnol.* 2019;4:25-35.
5. Oliveira R, Mendes J. Automação e eficiência na coleta de dados de patentes. *Rev Sist Intelig.* 2020;2:50-65.
6. Silva D, Ferreira P. Aplicações de web scraping na recuperação de dados científicos. *Rev Cienc Dados.* 2021;3:60-75.
7. Costa T, Ribeiro A. Redução do tempo na coleta de dados em bases públicas através de scraping. *Rev Tecnol Aplic.* 2022;1:45-55.
8. Almeida S. Interface de usuário e eficiência na navegação de grandes volumes de dados. *Rev Exp Usuário.* 2021;2:75-85.
9. Pereira H. Ferramentas de visualização e filtragem de dados para análise estruturada. *Inform Aplic.* 2020;1:50-60.
10. Machado F, Lopes M. Tecnologias modernas na degradação e processamento de informações. *Rev Bras Comput.* 2022;2:85-95.
11. Nogueira G. Desenvolvimento de interfaces responsivas para visualização de dados. *Interfaces Intelig.* 2021;1:100-110.
12. Santos A, Carvalho L. Ferramentas digitais na gestão do conhecimento em patentes. *Gest Conhecimento.* 2022;3:115-125.
13. Vieira M, Rocha E. O impacto da digitalização nos processos burocráticos da inovação. *Rev Bras Inov Tecnol.* 2023;4:85-95.
14. Silva FM, Santos RA, Silva RC. Patentscope (WIPO), Espacenet (EPO), BuscaWeb (INPI/BR): ferramentas de busca de patentes gratuitas. *Rev Tecnol Soc.* 2021;17(41):125-139. Available from: <https://periodicos.utfpr.edu.br/rts/article/view/14478>
15. Freitas L. Dashboards interativos para visualização de dados complexos. *Tecnol Visualização.* 2023;3:77-89.
16. Lima C. Integração de RPA em bases públicas. *Rev Bras Cienc Dados.* 2023;4:90-103.
17. Costa R. Automação e confiabilidade em sistemas de propriedade intelectual. *Rev Inov Digit.* 2023;5:40-52.
18. Oliveira R. Padronização e validação de dados em sistemas automatizados. *Rev Sist Intelig.* 2023;3:50-65.
19. Alvarez LO, Filgueiras R, Angeli R. Desafios na identificação de patentes em domínio público no Brasil. *Cad Prospec.* 2023;16(4):1292-1308. Available from: <https://www.revistas.ufba.br/index.php/nit/article/view/51958>