

**Preface**

|  |  |
| --- | --- |
| Project Title | Patent Analysis of Sustainable Aviation Fuel Using a Big Data Solution |
| ****Carried out by**** | Engineering students of the National School of Applied Sciences of Al Hoceima (ENSAH) ➢ Samiha El Mansouri ➢ Souhayla El Meftahi |
| ****Supervised by**** | Pr. Anas El Haddadi |
| ****Duration**** | From 1st March 2024 to 30th May 2024 |

**Abstract**

Patent analysis is an essential discipline for examining patent records on a global scale, aiming to identify crucial information and significant trends. Despite the availability of various proprietary software, there is still a lack of public, accessible, and user-friendly tools that promote open science and scientific software development.

In the context of the Ensah\_Aviation project, we have addressed this gap by designing and developing a public open-source tool. Leveraging patent data from various international sources, our tool is capable of generating descriptive analyses, thematic axes, and citation networks for patents worldwide. Its use and interpretation are demonstrated through a case study focusing on patents related to sustainable aviation, highlighting its functionalities and potential in this specific domain.

The results of our study demonstrate that Ensah\_Aviation is a practical resource for conducting patent analyses on a global scale, accessible even to individuals with limited experience in coding and software development.

**Keywords:** patent analysis; scientific software development;

topic modeling; citation networks

## **Introduction**

In the era of technological and entrepreneurial advancement, an increasing number of companies are seeking to safeguard their intellectual property. Particularly, the annual number of patent applications has nearly tripled over the past two decades, as per a study by the World Intellectual Property Organization (WIPO) ,making patent documents more valuable than ever. Patents are widely regarded as a secure choice for large corporations and organizations to protect their commercial rights, avoid legal disputes, and maintain their competitive edge .

The scale and significance of patenting become evident when considering the numerous patent offices worldwide responsible for receiving, evaluating, and granting patent applications. These offices, including prominent ones like the United States Patent and Trademark Office (USPTO), the European Patent Office (EPO), and the China National Intellectual Property Administration (CNIPA), undertake the challenging task of processing and analyzing patent documents, scrutinizing their objectives and validity. This wealth of information has led to the emergence of patent analysis (PA) as a promising scientific domain that leverages data from patent offices to extract valuable insights .

In essence, PA encompasses the study of patent documents using established methodologies and techniques such as text mining, machine learning, and data visualization .The insights derived from PA have myriad applications across various sections within organizations or businesses, including R&D management, human resources, mergers and acquisitions, business valuation, and competitive intelligence .Furthermore, PA offers ample opportunities for extracting meaningful information through the application of advanced approaches like topic modeling, network analysis, and machine learning.

However, despite the existence and utilization of PA tools, very few are available as free, accessible, and open-source solutions, with most either being proprietary or requiring payment after a short trial period. Moreover, existing open-source PA tools are somewhat complex to navigate, demanding a level of scientific knowledge. Consequently, the absence of a flexible, open-source, and publicly available PA tool catering to the needs of multiple user groups for research purposes constitutes a clear gap in the PA software domain.

Particularly in recent years, and especially amidst the COVID-19 pandemic, the programming community has strongly advocated for the principles of open science and the development of scientific software .These principles combine the need for transparency and openness across scientific domains with the creation of accessible software capable of processing and analyzing data using scientific concepts, thereby advancing science and primarily used for research.

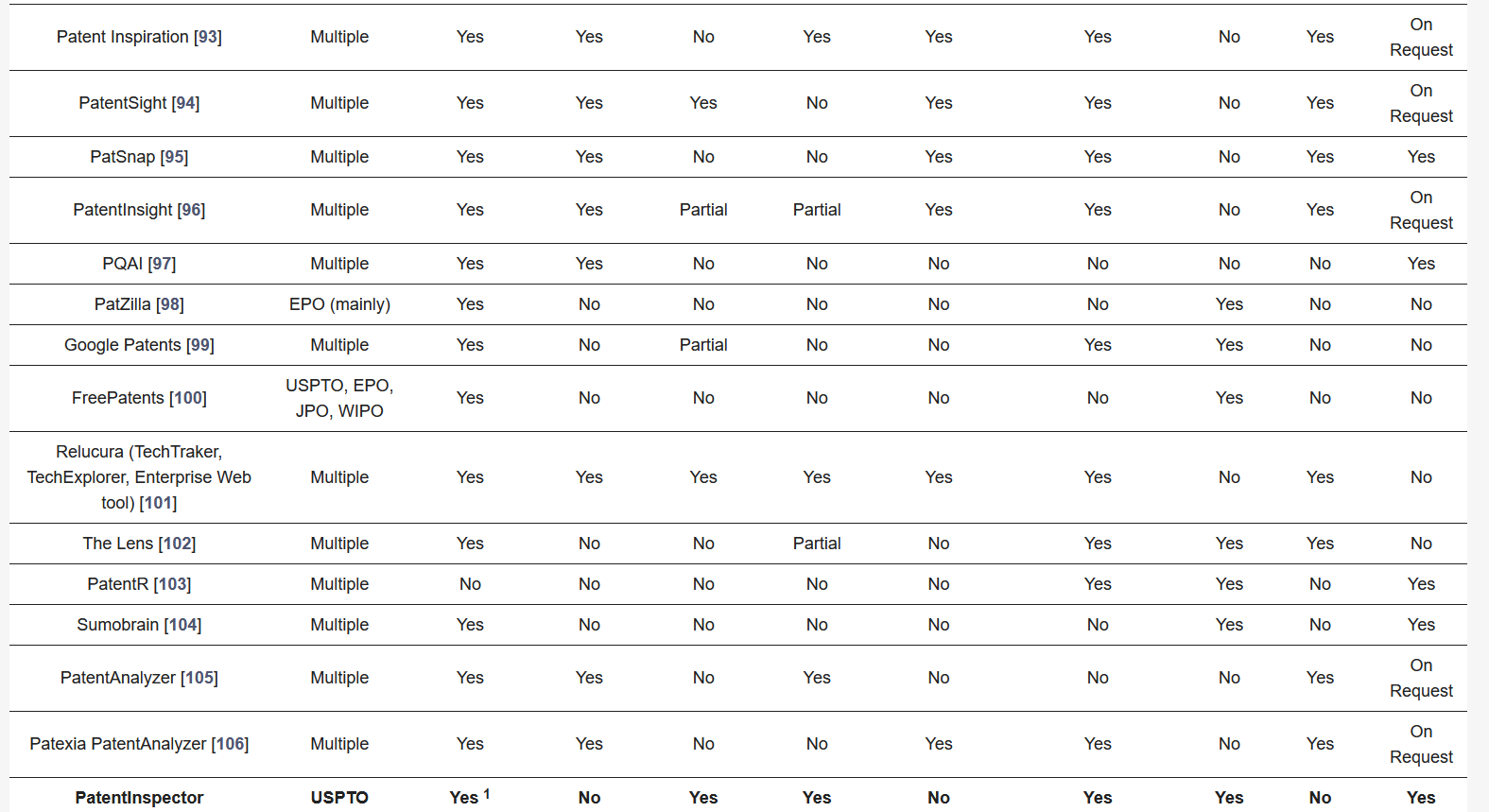
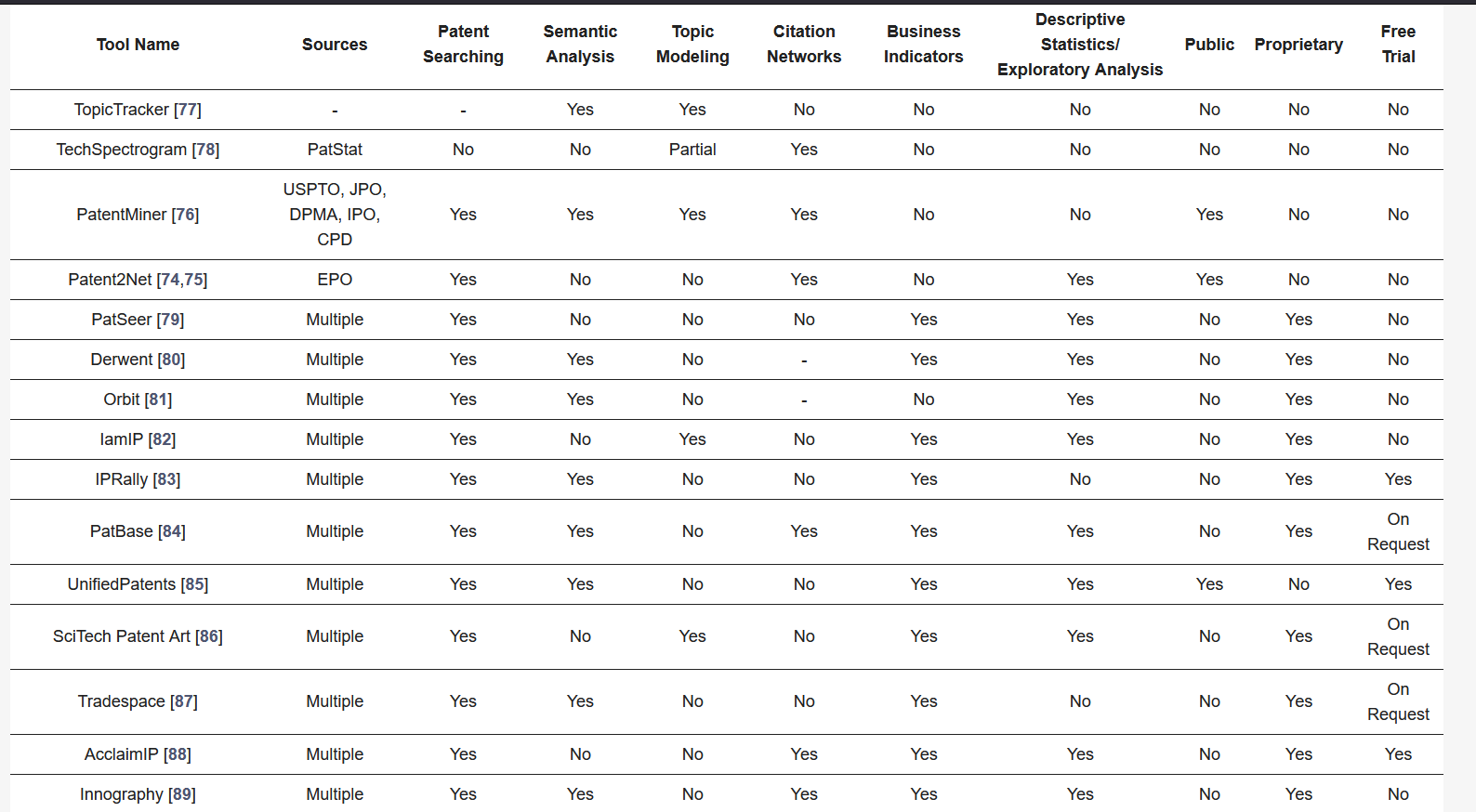
Acknowledging (a) the current lack of a public, intuitive, easy-to-use, and practical PA tool in contrast to numerous enterprise solutions, and (b) the growing movement for open science and the development of scientific applications opening new research avenues for scientists and practitioners unfamiliar with programming concepts, in this study, we introduce Ensah\_Aviation, an extensible open-source PA tool primarily implemented in Python and publicly deployed for wider usage.

Ensah\_Aviation recognizes the challenges associated with software deployment [11] and utilizes containers to mitigate them while providing a collective framework for patent record retrieval, processing, filtering, and analysis.

#### **Patent Analysis Tools**

As mentioned in the Introduction, there are several PA tools that allow the processing of patent records, widely used by enterprises and organizations. Table 1 presents basic information about the most popular PA tools, highlighting their key characteristics and operations. An inspection of the table reveals that the majority of the tools are proprietary and owned by large organizations (e.g., PatSeer, Derwent Innovation, Orbit Intelligence), with most of them providing access to millions of patent records from multiple offices. However, the fact that they are proprietary means that they do not offer a free trial (or may do so upon request) and typically require a subscription for their services. In addition, most of the proprietary tools focus on providing business indicators for patent growth (e.g., portfolio quality, investment value), often based on AI methodologies, while some also provide topic modeling or citation analysis functionalities.

Apart from proprietary PA tools, there are also several public tools that act as either PA suites or patent search databases. Among them, Patent2Net is an educational suite that leverages data from the EPO and focuses on citation networks and clustering. The suite also provides an interface that allows users to explore its capabilities and export results in various graph formats. The main target groups of Patent2Net are the educational and scientific communities ,while Ensah\_Aviation strives to include more target groups, such as industrial investors, developers, inexperienced researchers, and HR representatives. UnifiedPatents is another partially public PA suite that mainly focuses on business indicators and differs from Ensah\_Aviation, as it can be primarily used by business owners and economists. The portal provides an intuitive interface and companies with smaller revenue can use it for free, although it introduces a pricing option for larger companies. Finally, PatentMiner [76] is a notable effort undertaken before Ensah\_Aviation and provided an interface that executed advanced PA with topic modeling. The remaining free PA tools (PatZilla, FreePatents, and GooglePatents) are not typical PA tools, as they mainly provide advanced search engines for patent document retrieval. Thus, their PA capabilities are minimal and they cannot be considered similar to Ensah\_Aviation, which employs established scientific concepts and targets all types of users. GooglePatents ,in particular, stands as one of the most popular patent search engines, encompassing data from multiple patent offices and offering limited descriptive information (e.g., top inventors, top organizations).

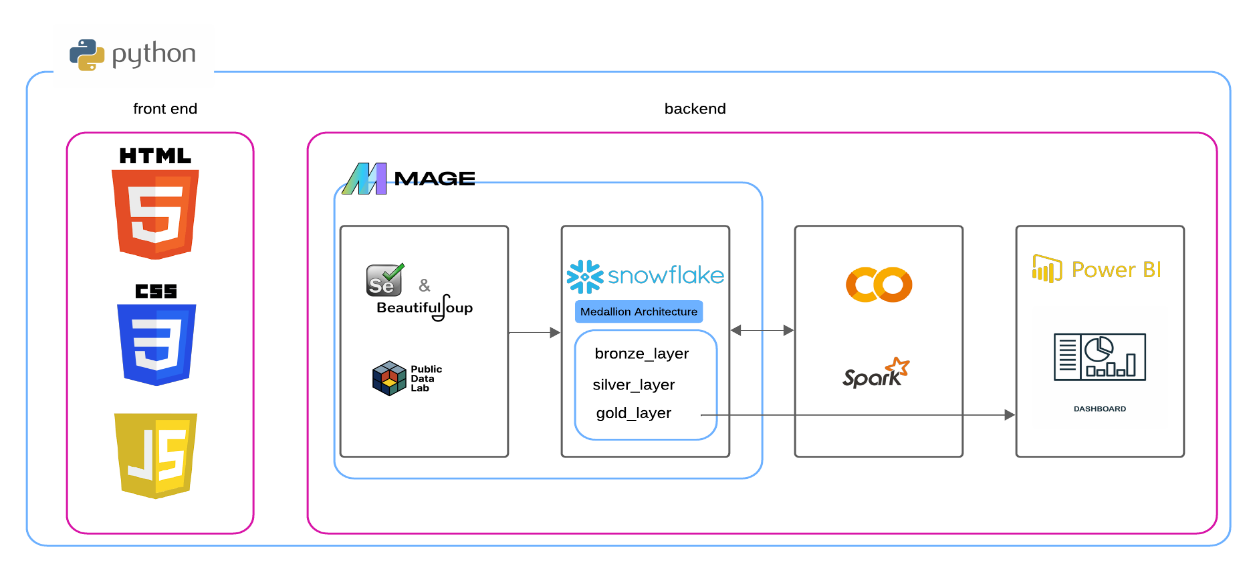


**Table 1.** Prominent patent analysis tools.

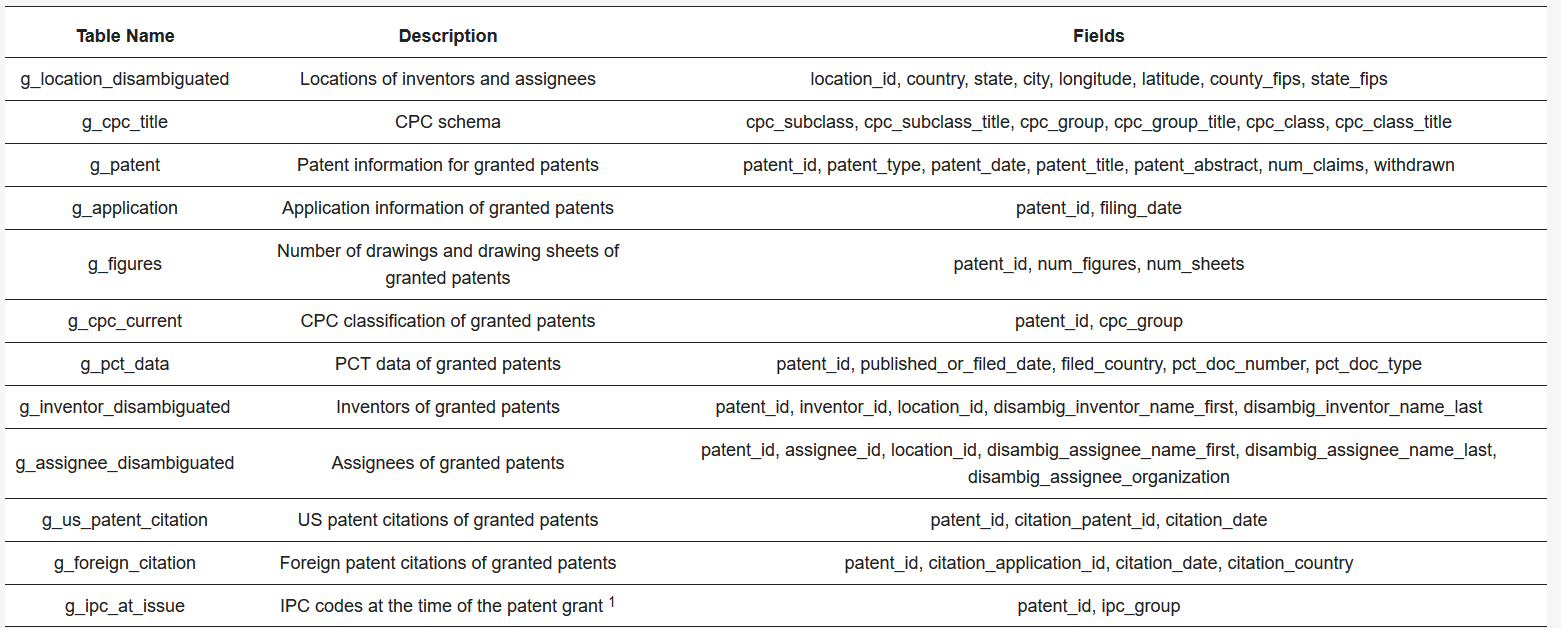
The analysis of PA tools and suites reveals that, as stated in the Introduction, while there is a plethora of such tools in the market and in software repositories, few of them are suitable for users with limited coding or scientific backgrounds. Ensah\_Aviation emerges to cover this deficit, with results from the USPTO while also offering different methodologies, efficient visualizations, and interpretable insights. In addition, Ensah\_Aviation introduces a novel perspective of PA for mainstream users and more advanced parties by including topic modeling methodologies that can profile the thematic axes of patent documents and aid users in making informed decisions.

## **Architecture and Workflow**

The backend of Ensah\_Aviation is developed in Python 3.11 to handle data storage, query administration, and application business logic. Complex operations are managed through SQL queries to interact with the database. Data analysis is conducted using Spark, which incorporates NLP tools for deeper understanding of patent documents. For the data infrastructure, we utilized Snowflake, employing a star schema for the data warehouse and an initial configuration as a data lake. Mage.ai have been integrated to facilitate data flow between Snowflake and Spark ,Power BI is employed for data visualization, enabling easy interpretation of analysis results. The frontend is built using a Bootstrap template, providing a user-friendly and responsive interface to interact with the application's features.

**Data Collection, Preprocessing and Storage**

Ensah\_Aviation operates on patent record data provided free of charge by the USPTO. Specifically, the database used relies on bulk data available on the USPTO's PatentsView platform, which serves as a repository for all patents registered and granted by the USPTO and is regularly updated. Additionally, we expanded our data sources to include resources such as Google Patent Lens and UPTO Net Space. To enrich our dataset, we also implemented web scraping techniques using tools such as BeautifulSoup and selenium. The tool incorporates a management utility named "USPTO," depicted in Figure 3, which automates the process of downloading, decompressing, preprocessing, and inserting data into the database. This database is structured into tables, each containing different aspects of patent records such as patent classes, patent inventors, etc. Additionally, we utilized Beautiful Soup and selenium for scraping data from various APIs

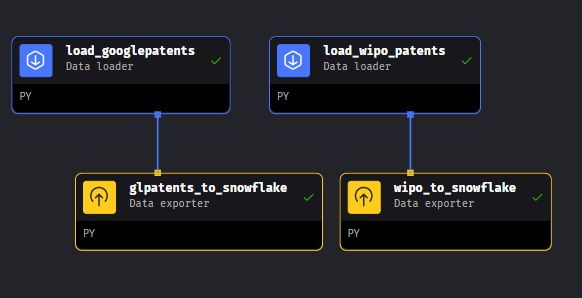


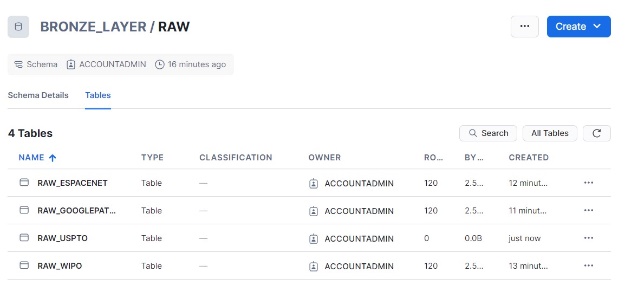
**Table 2. Fields retrieved from the USPTO**

After scraping the data, we established an automation process to integrate it into Snowflake using Mage. To ensure efficient data management, we adopted a three-tiered storage approach: bronze, silver, and gold.

In the Bronze layer, we store raw data as collected after the initial scraping. This data is kept in its original form without any preprocessing, serving as a repository for unfiltered and untreated information. Next, the Silver layer comes into play. After scraping, the data undergoes preprocessing to remove outliers, duplicates, and errors. The cleaned and preprocessed data is then stored in the Silver layer. This layer acts as an intermediary stage where the data is ready for deeper analysis but has not yet been transformed into actionable insights. Finally, in the Gold layer, the preprocessed data from the Silver layer is transformed and aggregated to meet the specific needs of end users. The data is organized into a star schema to facilitate queries and analysis. This layer contains data ready for use in report generation, business analysis, and other business applications. To start, we automated the scraping process using Mage and stored the data in the Bronze layer in Snowflake.

First, we started by automating the scraping process using Mage and storing the data in the bronze layer in Snowflake.





**Figure 3: representing the scraping pipeline using Mage into Snowflake**

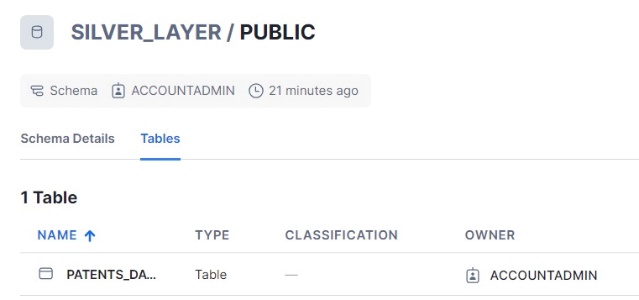
After downloading the tables of interest, containing information about approximately 220 000 patents, an automated preprocessing procedure is initiated. The preprocessing deployed in Ensah\_Aviation involves stop word removal and lemmatization of text fields such as the patent title or abstract, to facilitate and expedite the text analysis conducted in subsequent stages. Additionally, in this phase, computations are performed in advance for optimization purposes and stored as additional columns, effectively constructing a sort of long-term database cache. Table 3 summarizes all the precomputed fields that contribute to increasing the throughput of the application

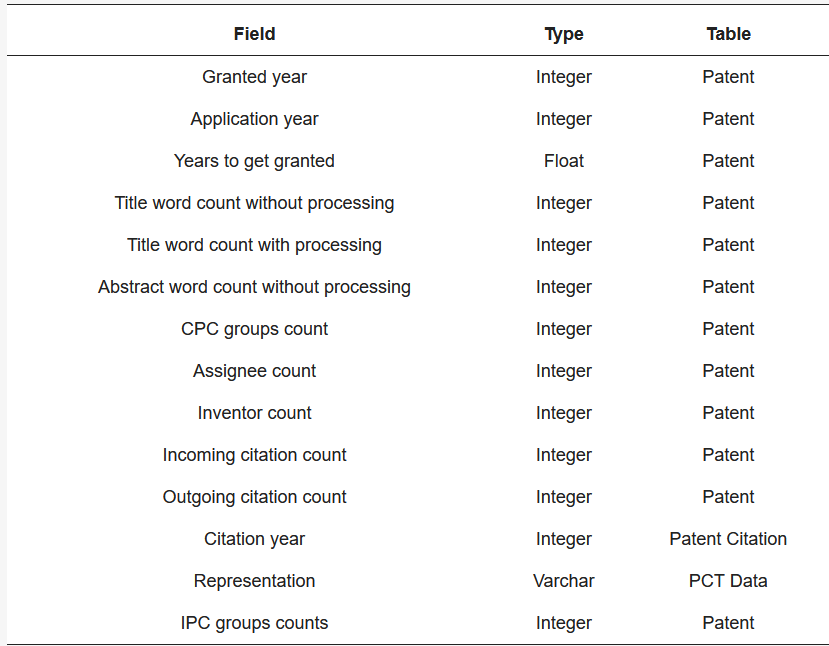
Merge all tables in bronze layer to globale table which containes the columns we will use in analyse

* **Silver Layer:**

In the silver layer, we stored the cleaned data, eliminating duplicates and removing data originating from CN due to difficulties encountered with UTF-8 encoding of these records

In the Silver layer, we remove columns that are not present in all sources (Google Patents, USPTO, WIPO, Espacenet). We retain only the common columns to store them in the Gold layer, organized in a star schema.





**Table 3. Precomputed database fields of Ensah\_Aviation from USPTO**

#### **Computation**

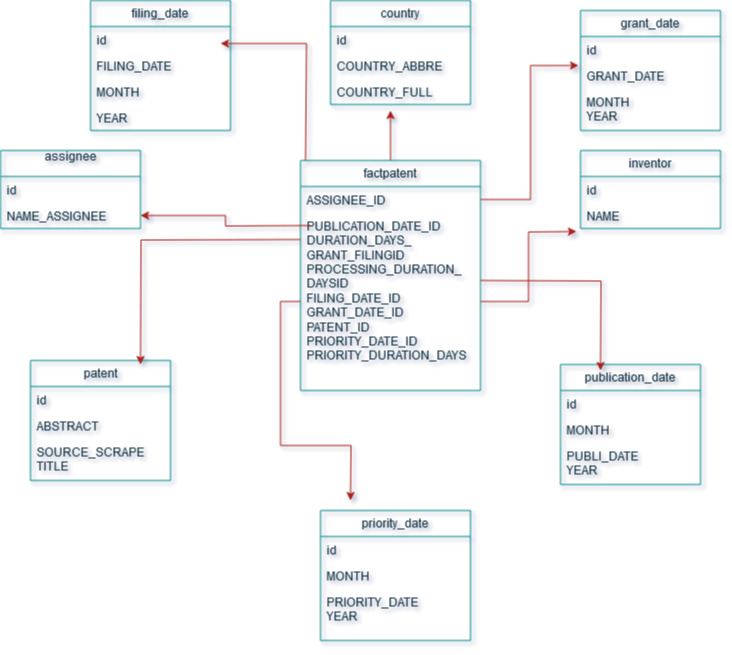
In the computation phase, we added a column containing the country information. We noticed that the ID column includes both the ID of each patent and the country abbreviation. For Google Patent data specifically, we extracted the abbreviation and transformed it into the full country name, which we then stored in a new column named "country." However, for other sources such as USPTO, Espacenet, and WIPO, the country information was already available, so no extraction was needed, and the country column was populated directly from existing data.

Additionally, we created three calculated columns by computing the difference between specific existing columns: "filing\_to\_priority\_duration\_days" captures the difference between the "filing/creation date" and the "priority date," measuring the time from when a patent is first filed to when it claims priority. This duration is crucial for understanding the delay between the initial filing and the establishment of priority, impacting the patent's legal standing and competitive edge. "creation\_to\_publication\_duration\_days" calculates the difference between the "publication date" and the "creation date," reflecting the time taken for a patent to be published after its creation. This metric is important for tracking how long it takes for a patent to become publicly accessible, which is vital for transparency and for informing potential competitors or collaborators. "grant\_to\_filing\_duration\_days" records the difference between the "grant date" and the "filing/creation date," indicating the duration from when a patent is filed to when it is granted. This period represents the time taken for the patent office to review and approve the patent application, affecting the patent's enforceability and the inventor's ability to capitalize on the innovation. These calculated columns measure the durations between key stages in the patent process.

For the "technology\_sector" column addition, we chose to implement a Support Vector Machine (SVM) classifier due to its effectiveness in handling high-dimensional data and its capability to handle non-linear relationships between features. The SVM algorithm was utilized as follows: During data preparation, the abstract and title text data from the patent records were preprocessed, including tokenization, stopword removal, and lemmatization. Feature extraction involved applying TF-IDF transformation to convert the preprocessed text data into numerical feature vectors representing the abstracts and titles in a high-dimensional space, with each dimension corresponding to a unique term in the corpus. The model training phase involved training the SVM model using the TF-IDF vectors of the abstracts and titles as input features, with each patent record labeled with its corresponding technology sector. The SVM algorithm learned the decision boundary that best separates the different technology sectors in the feature space. Model evaluation was conducted using validation techniques such as cross-validation, and metrics like accuracy, precision, recall, and F1-score were calculated to assess the model's effectiveness in classifying patents into the correct technology sectors. For prediction, the trained SVM model could predict the technology sector of new patents based on their abstracts and titles, assigning each new patent to the most probable technology sector based on the learned decision boundary.

After preprocessing, the entire set of patent records was inserted into CSV files. We created several files, including "factpatent," "assignee," "country," "filing\_date," "grant\_date," "inventor," "patent," "priority\_date," and "publication\_date." These files were then imported into our data warehouse schema in Snowflake. This approach addresses the issue of distributed data from different sources (USPTO, Google Patent, Espacenet, WIPO) by retaining only the common data.

.**Gold Layer:**

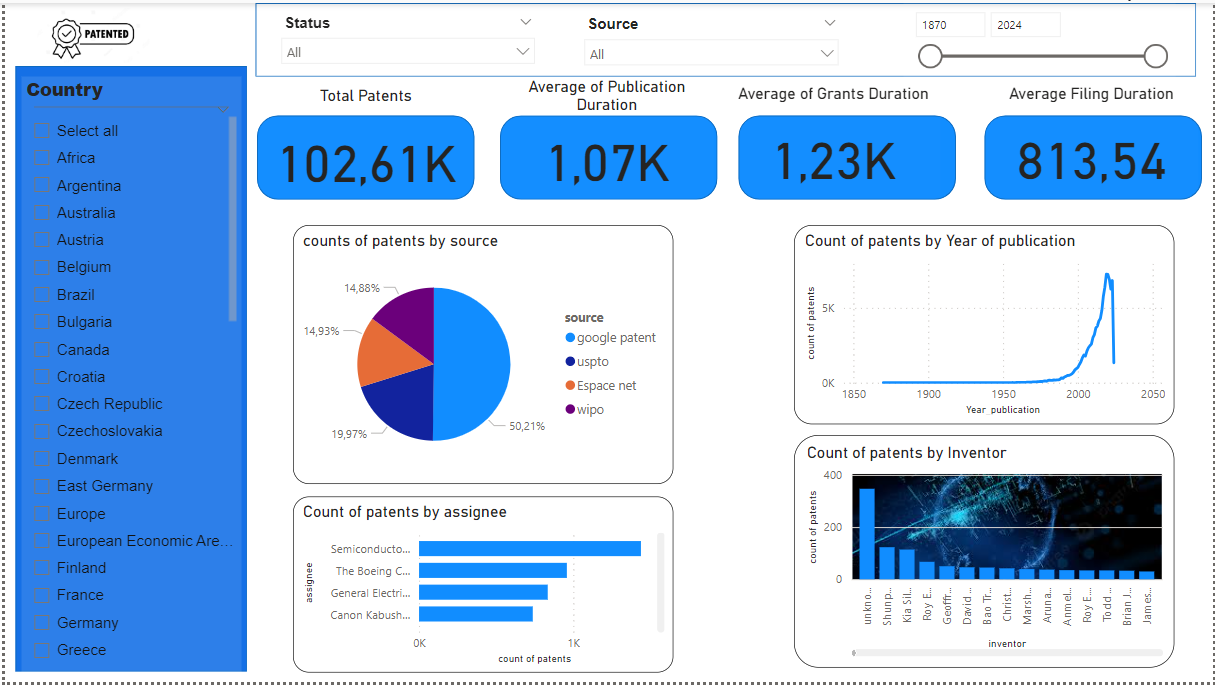


**Figure 4.** The ER diagram of the computational-related tables.

**Visualization**

During this analysis, we provided a comprehensive view of patent data. We calculated the total and average durations of patent publication, grant, and filing periods, thereby offering a comprehensive overview of these crucial metrics spanning from 1870 to 2024. The dynamic aspect of this analysis lies in its ability to adapt to user selections, whether it's choosing the type of patent source or selecting a specific country.

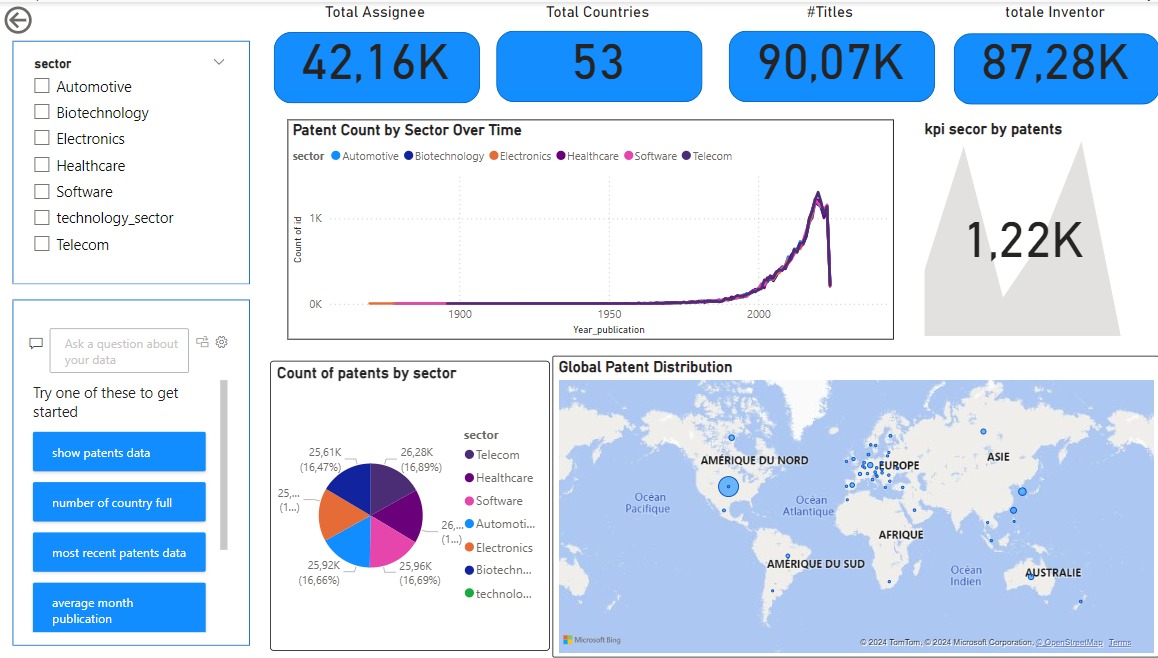
Furthermore, we furnished detailed data on patent distribution across various dimensions. This encompassed the number of patents by source, allowing for an understanding of patent distribution based on their origin. We also provided insights into patent distribution by assignee, highlighting key players in the innovation landscape. Additionally, we presented a temporal analysis by displaying the number of patents published each year, enabling the identification of trends and fluctuations in the intellectual property domain. Lastly, we shed light on the significance of inventors by furnishing statistics on the number of patents per inventor, providing an insight into individual contributions to innovation.

****

**Figure 1: Patent Analysis**

**Analysis Page 2 Description**

Continuing from the comprehensive analysis of patent data across various sectors, Page 2 delves into more detailed and specific insights, providing a deeper understanding of patent distribution and trends. This page includes several key sections. The top section, Sector-Specific Trends, features line or bar charts representing sectors such as Automotive, Biotechnology, Electronics, Healthcare, Software, and Telecom, highlighting year-over-year changes in patent filings to illustrate innovation peaks and declines. The middle section, Patent Distribution by Country, presents a heat map or bar chart illustrating the distribution of patents by country, emphasizing global leaders in patent filings. The bottom left section, Top Assignees and Inventors, displays a table or bar chart of the top patent holders and inventors, offering insights into key contributors to technological development. The bottom right section, Sectoral KPI Analysis, presents detailed Key Performance Indicators (KPIs) such as average time to grant, patent citation index, and patent renewal rates, providing a deeper understanding of the quality and impact of patents. Additionally, the left panel features interactive filters and queries, allowing users to refine data by sub-sectors, patent status, and other criteria, along with an advanced query box offering suggestions for in-depth analysis. The visual elements include sector-specific trend charts with clear legends and annotations, a heat map with color gradients or sorted bar charts for patent distribution by country, tables and bar charts for top assignees and inventors, KPI metrics presented using gauges or charts, and a user-friendly filter panel with interactive options and an autocomplete query box.



**Analysis Page 3 Description**

Page 3 of the analysis introduces the Patent Process Efficiency Analysis dashboard. This dashboard provides insights into the variability of patent filing and approval times.

The Long Grant Duration section indicates that a lengthy grant process suggests various factors such as thorough examination, administrative delays, or workload. The Creation to Publication Duration Similar to Grant Duration section suggests that much of the time from filing to grant is spent preparing the patent for publication. Shorter Processing Duration section implies that once the patent application enters the processing phase, it doesn't take as long to process compared to the total time from filing to grant. This suggests that initial stages before detailed examination and final steps after approval consume significant time.

The Understanding the Process section highlights that the longest part of the patent process is the time it takes for a patent to be granted after filing. This includes both the initial review and final approval steps.

Efficiency Insights suggest that most of the work occurs before the patent is officially published, as indicated by the similar durations from creation to publication and total time to grant.

Areas for Improvement include understanding the causes of delays before and after processing to improve the overall timeline.

Moving to the Patent Process Variability dashboard, the Condensed Center Points section suggests that the majority of patents have similar filing-to-priority and grant-to-filing durations, indicating fairly consistent process times. Outliers, represented by larger or smaller points, may require further investigation to understand significant deviations from the majority.

The analysis suggests using filters to isolate outliers and analyze their characteristics, such as specific assignees, countries, or patent types.

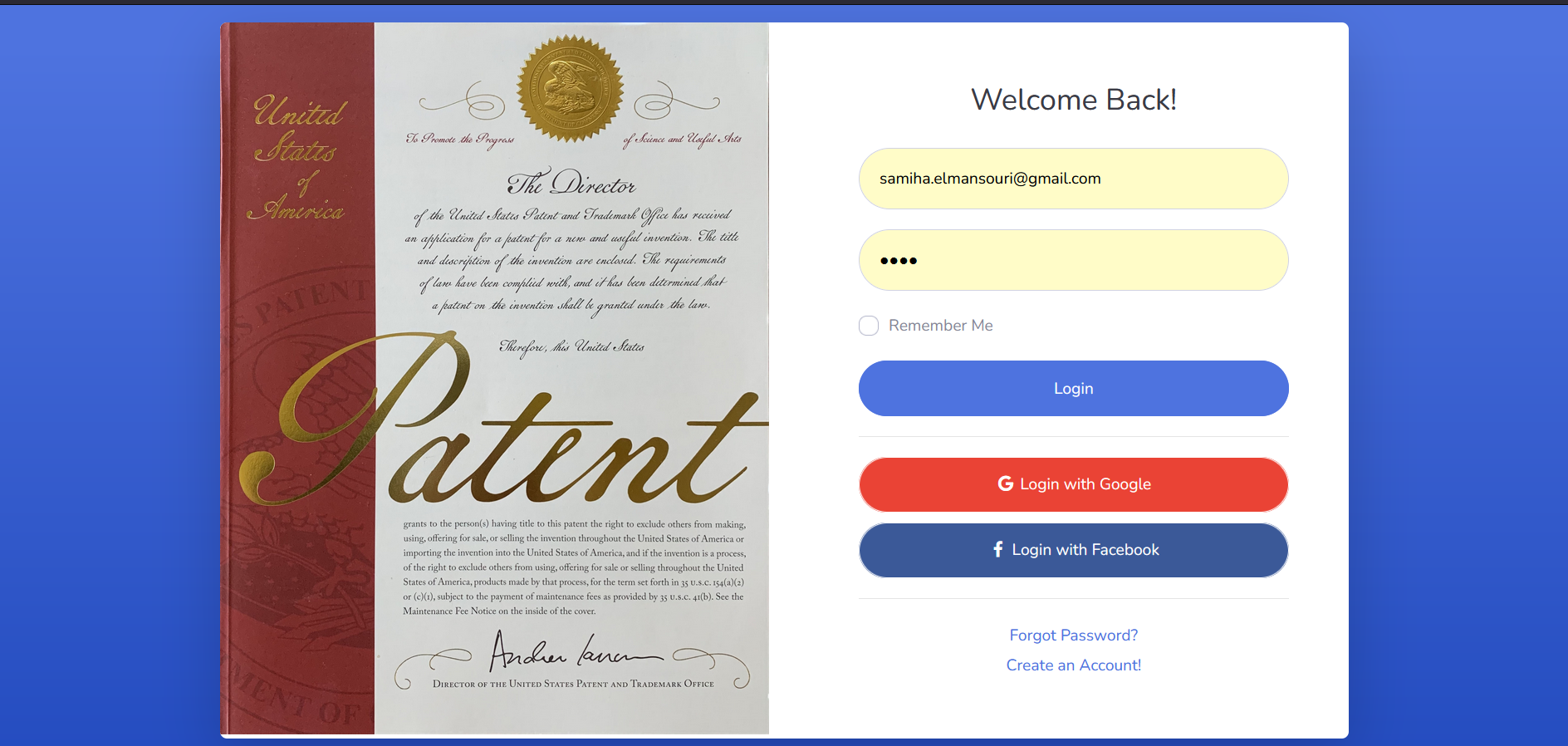
Interpreting the horizontal and vertical distributions, the consistent grant duration and variable filing duration indicate operational efficiency in grant processes but potential improvement opportunities in filing processes.

The Key Influencers visualization provides insights into factors influencing grant durations, such as country, inventor, or sector. These insights can inform optimization strategies, including addressing inefficiencies, promoting innovation, and fostering collaboration and knowledge sharing.

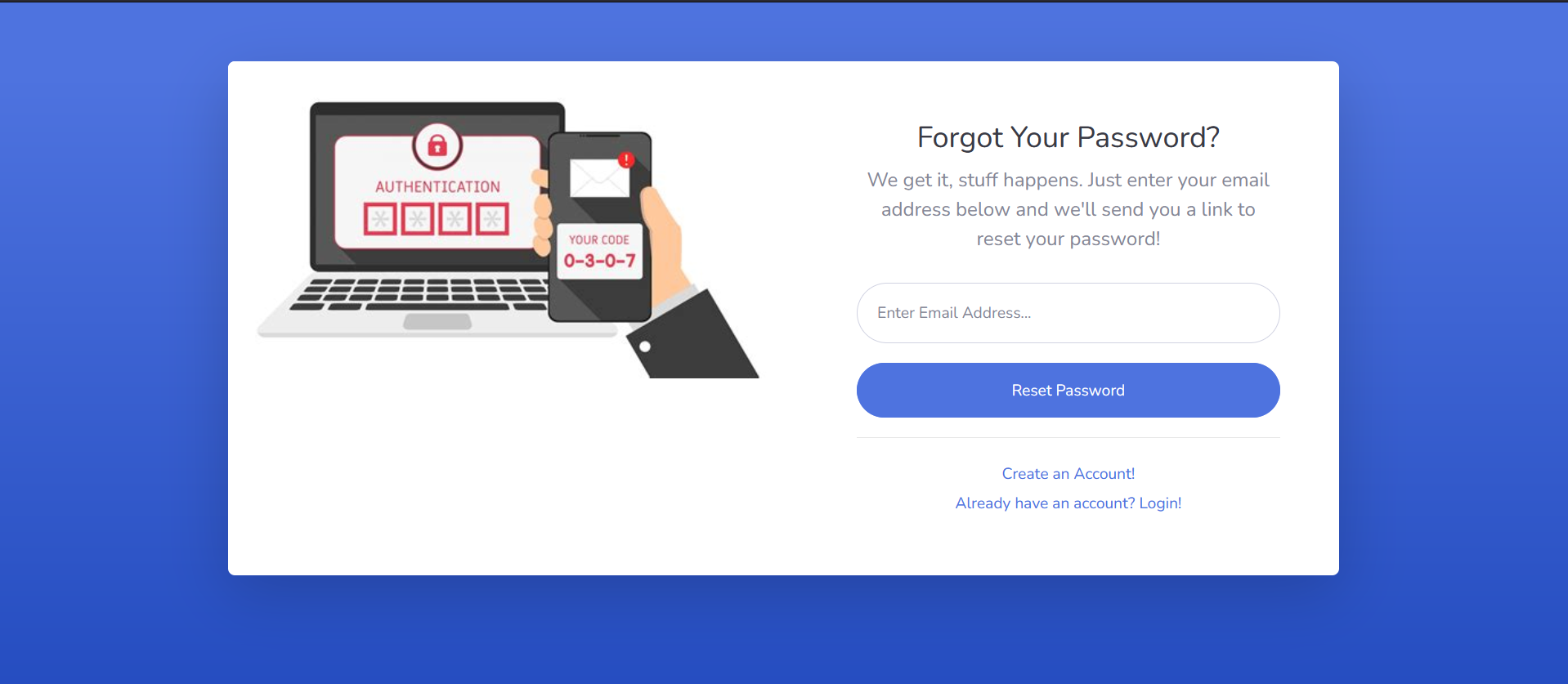
In summary, the analysis identifies areas of efficiency and potential improvement in the patent process, offering actionable insights to streamline operations and drive innovation.



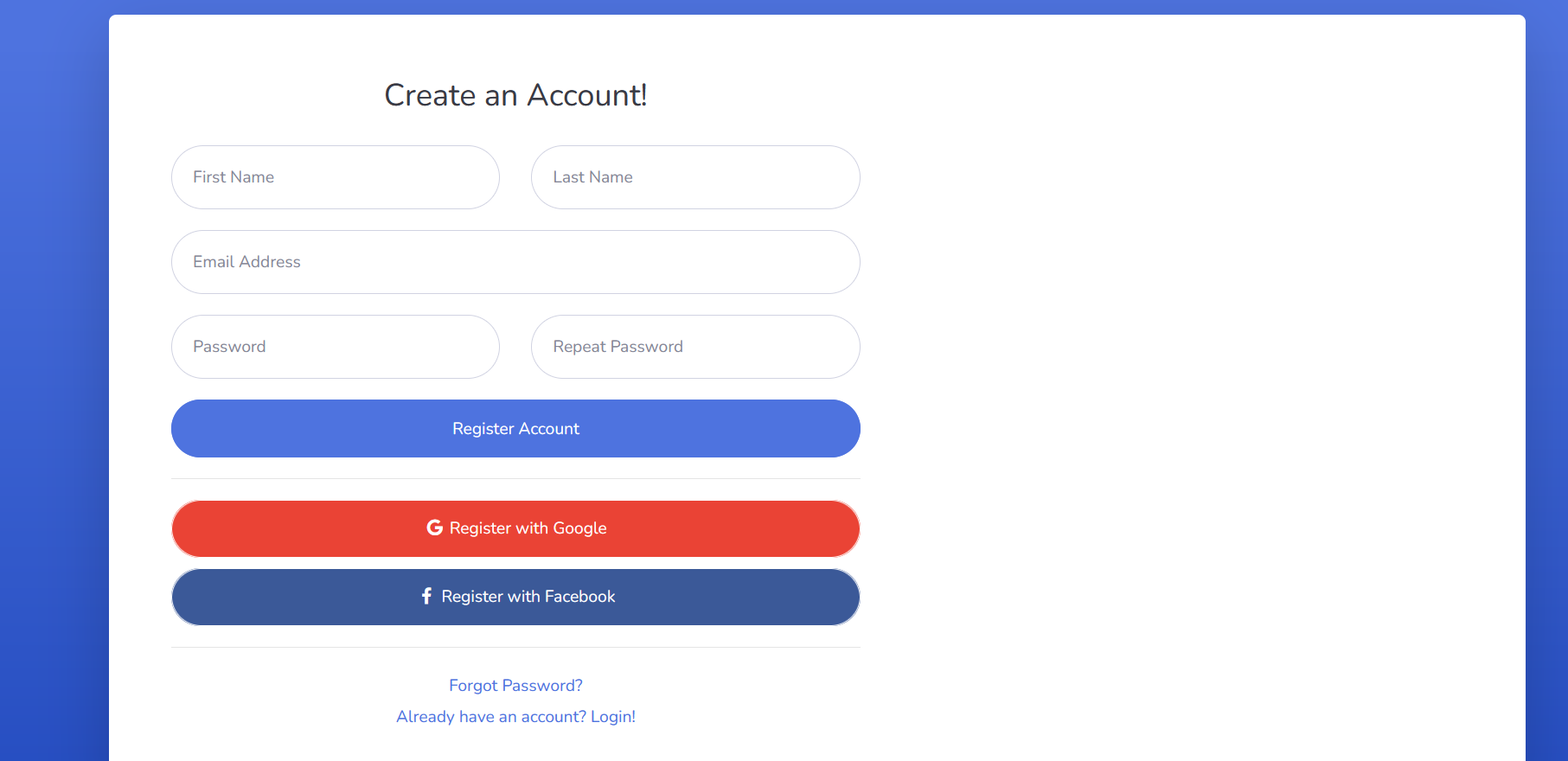
**appendix page**



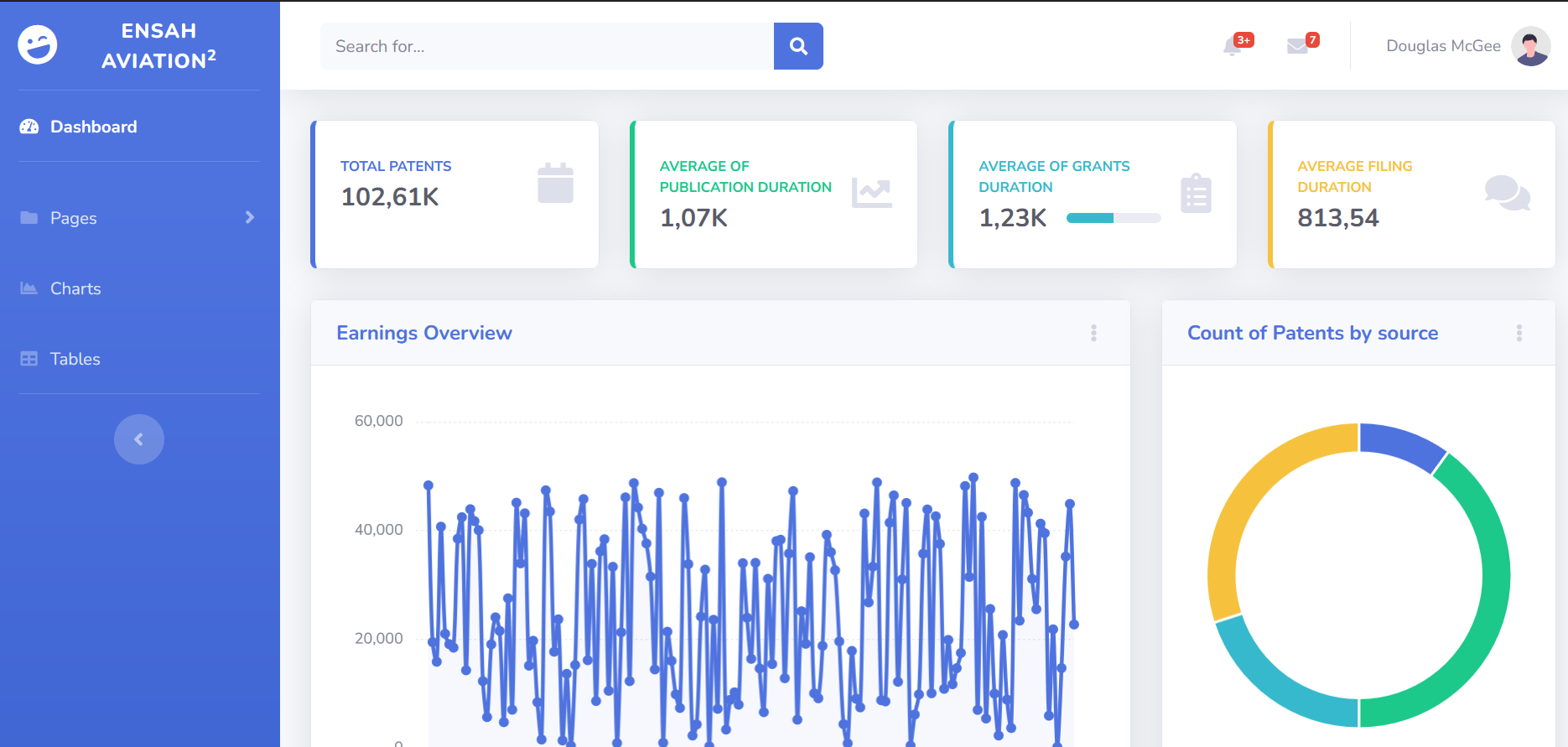
**Figure 1 :Login page**

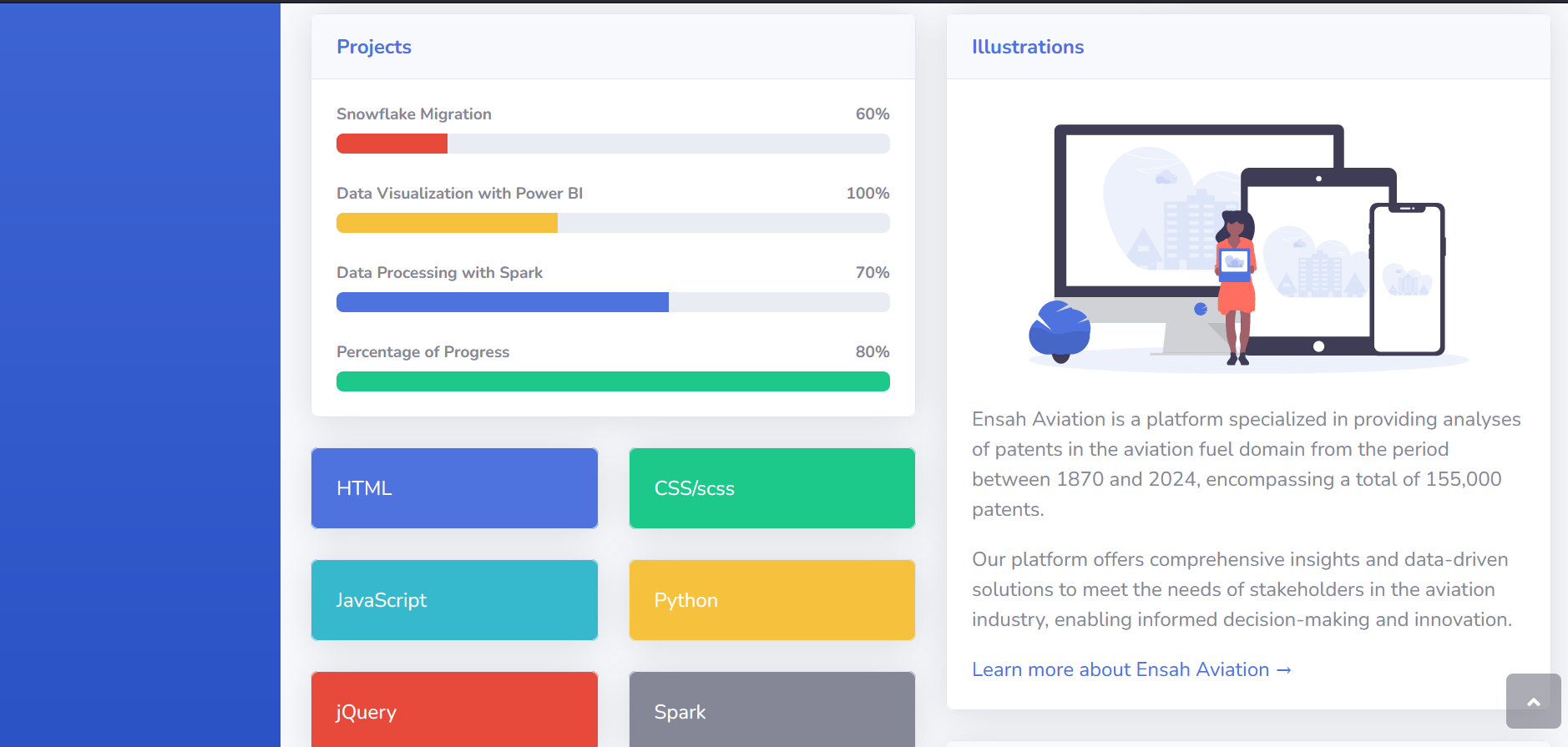


**Forgot password page**

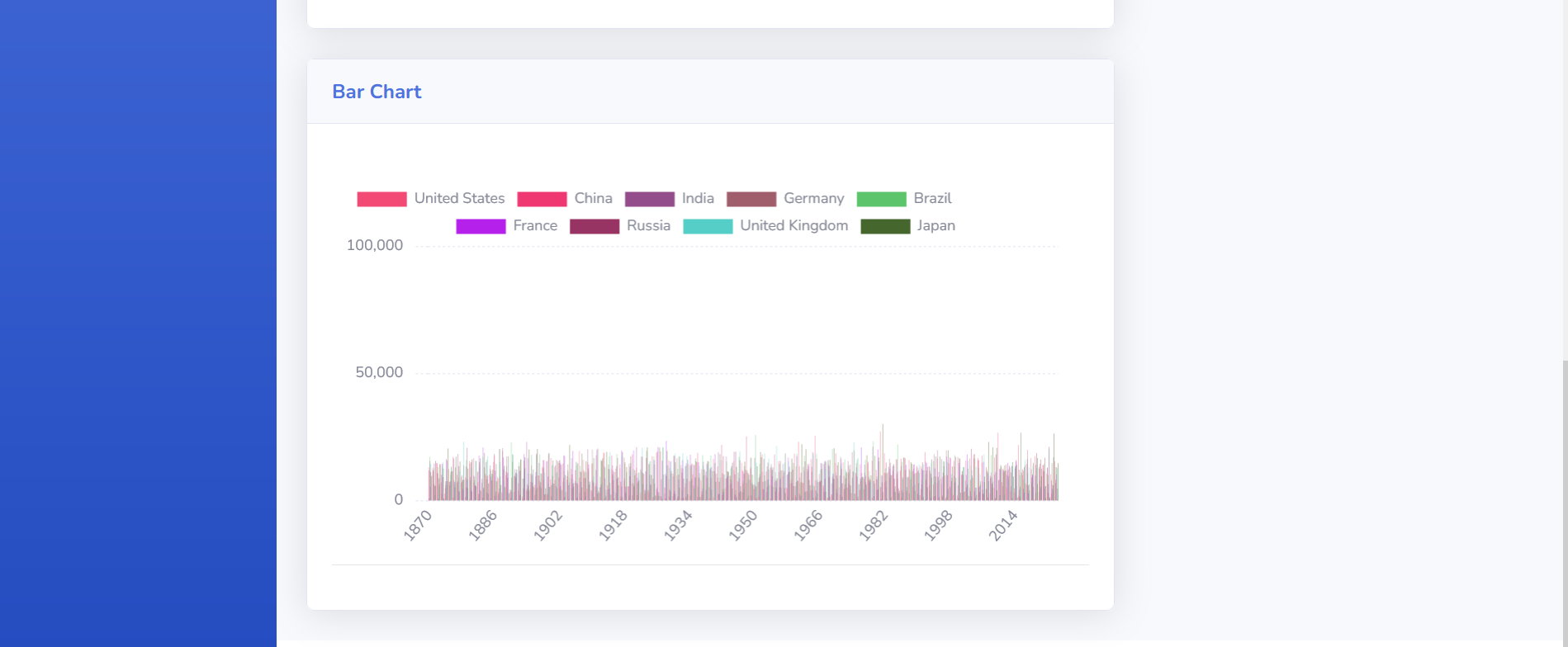


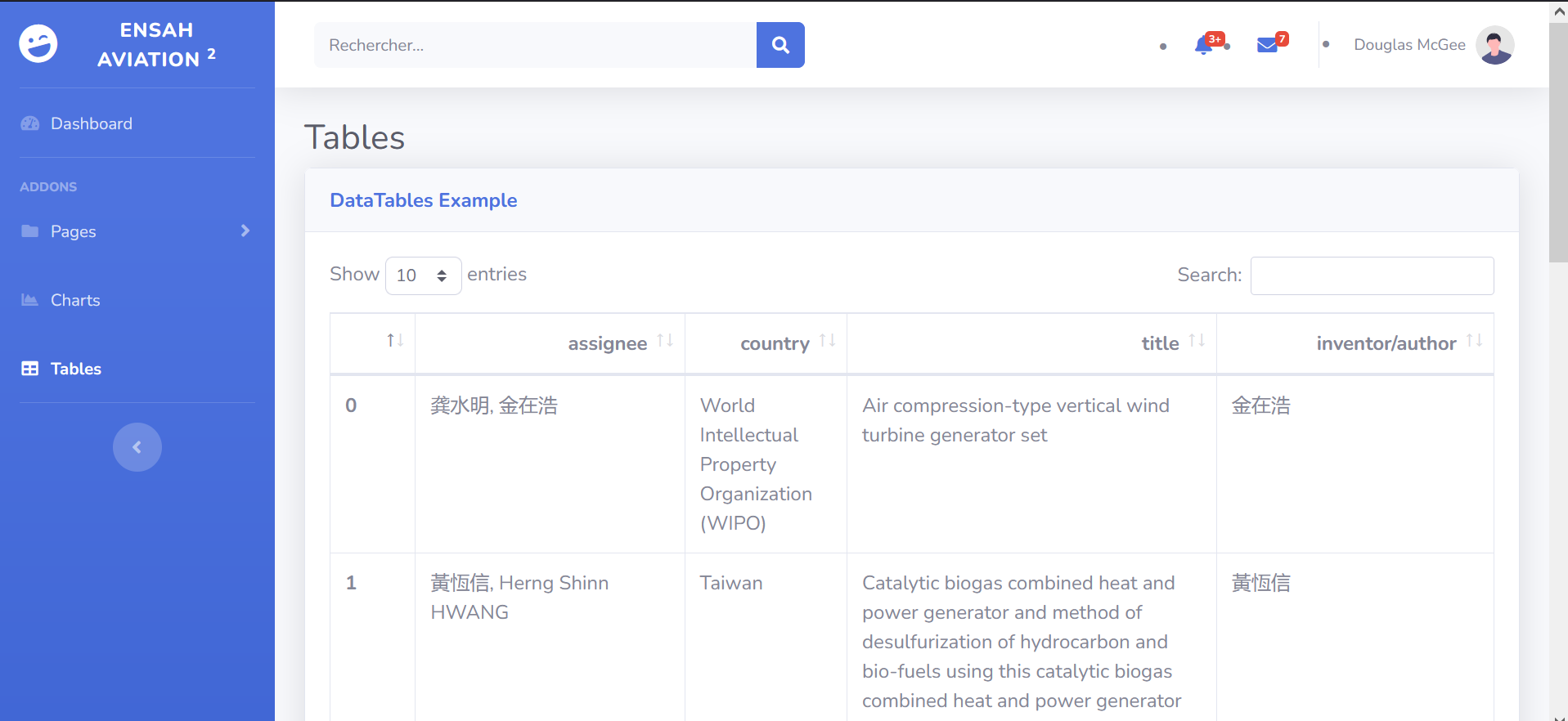
**Register page**



**Dashboard**







## **Table search page**

## **Chart page**

## **References**

 Dyvik, E.H. Number of Patent Applications Worldwide 2021. Available online: <https://www.statista.com/statistics/257610/number-of-patent-applications-worldwide> (accessed on 29 October 2023).

 Lemley, M. Understanding the realities of modern patent litigation [Preprint]. Tex. Law Rev. **2016**, 92, 1769. [[Google Scholar](https://scholar.google.com/scholar_lookup?title=Understanding+the+realities+of+modern+patent+litigation+%5BPreprint%5D&author=Lemley,+M.&publication_year=2016&journal=Tex.+Law+Rev.&volume=92&pages=1769&doi=10.31235/osf.io/ascgv)] [[CrossRef](https://doi.org/10.31235/osf.io/ascgv" \t "_blank)]

 Abraham, B.P.; Moitra, S.D. Innovation assessment through patent analysis. Technovation **2001**, 21, 245–252. [[Google Scholar](https://scholar.google.com/scholar_lookup?title=Innovation+assessment+through+patent+analysis&author=Abraham,+B.P.&author=Moitra,+S.D.&publication_year=2001&journal=Technovation&volume=21&pages=245%E2%80%93252&doi=10.1016/S0166-4972(00)00040-7)] [[CrossRef](https://doi.org/10.1016/S0166-4972(00)00040-7" \t "_blank)]

 Abbas, A.; Zhang, L.; Khan, S.U. A literature review on the state-of-the-art in patent analysis. World Pat. Inf. **2014**, 37, 3–13. [[Google Scholar](https://scholar.google.com/scholar_lookup?title=A+literature+review+on+the+state-of-the-art+in+patent+analysis&author=Abbas,+A.&author=Zhang,+L.&author=Khan,+S.U.&publication_year=2014&journal=World+Pat.+Inf.&volume=37&pages=3%E2%80%9313&doi=10.1016/j.wpi.2013.12.006)] [[CrossRef](https://doi.org/10.1016/j.wpi.2013.12.006" \t "_blank)]

 Tseng, Y.-H.; Lin, C.-J.; Lin, Y.-I. Text mining techniques for patent analysis. Inf. Process. Manag. **2007**, 43, 1216–1247. [[Google Scholar](https://scholar.google.com/scholar_lookup?title=Text+mining+techniques+for+patent+analysis&author=Tseng,+Y.-H.&author=Lin,+C.-J.&author=Lin,+Y.-I.&publication_year=2007&journal=Inf.+Process.+Manag.&volume=43&pages=1216%E2%80%931247&doi=10.1016/j.ipm.2006.11.011)] [[CrossRef](https://doi.org/10.1016/j.ipm.2006.11.011" \t "_blank)]

 Breitzman, A.F.; Mogee, M.E. The many applications of patent analysis. J. Inf. Sci. **2002**, 28, 187–205. [[Google Scholar](https://scholar.google.com/scholar_lookup?title=The+many+applications+of+patent+analysis&author=Breitzman,+A.F.&author=Mogee,+M.E.&publication_year=2002&journal=J.+Inf.+Sci.&volume=28&pages=187%E2%80%93205&doi=10.1177/016555150202800302)] [[CrossRef](https://doi.org/10.1177/016555150202800302" \t "_blank)]

 Foster, E.D.; Deardorff, A. Open science framework (OSF). J. Med. Libr. Assoc. JMLA **2017**, 105, 203. [[Google Scholar](https://scholar.google.com/scholar_lookup?title=Open+science+framework+(OSF)&author=Foster,+E.D.&author=Deardorff,+A.&publication_year=2017&journal=J.+Med.+Libr.+Assoc.+JMLA&volume=105&pages=203&doi=10.5195/jmla.2017.88)] [[CrossRef](https://doi.org/10.5195/jmla.2017.88" \t "_blank)]

 Vicente-Saez, R.; Martinez-Fuentes, C. Open Science now: A systematic literature review for an integrated definition. J. Bus. Res. **2018**, 88, 428–436. [[Google Scholar](https://scholar.google.com/scholar_lookup?title=Open+Science+now:+A+systematic+literature+review+for+an+integrated+definition&author=Vicente-Saez,+R.&author=Martinez-Fuentes,+C.&publication_year=2018&journal=J.+Bus.+Res.&volume=88&pages=428%E2%80%93436&doi=10.1016/j.jbusres.2017.12.043)] [[CrossRef](https://doi.org/10.1016/j.jbusres.2017.12.043" \t "_blank)]

 Segal, J.; Morris, C. Developing Scientific Software. IEEE Softw. **2008**, 25, 18–20. [[Google Scholar](https://scholar.google.com/scholar_lookup?title=Developing+Scientific+Software&author=Segal,+J.&author=Morris,+C.&publication_year=2008&journal=IEEE+Softw.&volume=25&pages=18%E2%80%9320&doi=10.1109/MS.2008.85)] [[CrossRef](https://doi.org/10.1109/MS.2008.85" \t "_blank)]

 Nguyen-Hoan, L.; Flint, S.; Sankaranarayana, R. A survey of scientific software development. In Proceedings of the 2010 ACM-IEEE International Symposium on Empirical Software Engineering and Measurement, Bolzano/Bozen, Italy, 16–17 September 2010; pp. 1–10. [[Google Scholar](https://scholar.google.com/scholar_lookup?title=A+survey+of+scientific+software+development&conference=Proceedings+of+the+2010+ACM-IEEE+International+Symposium+on+Empirical+Software+Engineering+and+Measurement&author=Nguyen-Hoan,+L.&author=Flint,+S.&author=Sankaranarayana,+R.&publication_year=2010&pages=1%E2%80%9310)]

 Arcangeli, J.-P.; Boujbel, R.; Leriche, S. Automatic deployment of distributed software systems: Definitions and state of the art. J. Syst. Softw. **2015**, 103, 198–218. [[Google Scholar](https://scholar.google.com/scholar_lookup?title=Automatic+deployment+of+distributed+software+systems:+Definitions+and+state+of+the+art&author=Arcangeli,+J.-P.&author=Boujbel,+R.&author=Leriche,+S.&publication_year=2015&journal=J.+Syst.+Softw.&volume=103&pages=198%E2%80%93218&doi=10.1016/j.jss.2015.01.040)] [[CrossRef](https://doi.org/10.1016/j.jss.2015.01.040" \t "_blank)]

 Available online: <https://github.com/search?q=patent+analysis&type=repositories> (accessed on 29 October 2023).

 Ardito, L.; D’Adda, D.; Petruzzelli, A.M. Mapping innovation dynamics in the Internet of Things domain: Evidence from patent analysis. Technol. Forecast. Soc. Chang. **2018**, 136, 317–330. [[Google Scholar](https://scholar.google.com/scholar_lookup?title=Mapping+innovation+dynamics+in+the+Internet+of+Things+domain:+Evidence+from+patent+analysis&author=Ardito,+L.&author=D%E2%80%99Adda,+D.&author=Petruzzelli,+A.M.&publication_year=2018&journal=Technol.+Forecast.+Soc.+Chang.&volume=136&pages=317%E2%80%93330&doi=10.1016/j.techfore.2017.04.022)] [[CrossRef](https://doi.org/10.1016/j.techfore.2017.04.022" \t "_blank)]

 Fujii, H.; Managi, S. Trends and priority shifts in artificial intelligence technology invention: A global patent analysis. Econ. Anal. Policy **2018**, 58, 60–69. [[Google Scholar](https://scholar.google.com/scholar_lookup?title=Trends+and+priority+shifts+in+artificial+intelligence+technology+invention:+A+global+patent+analysis&author=Fujii,+H.&author=Managi,+S.&publication_year=2018&journal=Econ.+Anal.+Policy&volume=58&pages=60%E2%80%9369&doi=10.1016/j.eap.2017.12.006)] [[CrossRef](https://doi.org/10.1016/j.eap.2017.12.006" \t "_blank)]

 Tseng, C.-Y.; Ting, P.-H. Patent analysis for technology development of artificial intelligence: A country-level comparative study. Innovation **2013**, 15, 463–475. [[Google Scholar](https://scholar.google.com/scholar_lookup?title=Patent+analysis+for+technology+development+of+artificial+intelligence:+A+country-level+comparative+study&author=Tseng,+C.-Y.&author=Ting,+P.-H.&publication_year=2013&journal=Innovation&volume=15&pages=463%E2%80%93475&doi=10.5172/impp.2013.15.4.463)] [[CrossRef](https://doi.org/10.5172/impp.2013.15.4.463" \t "_blank)]

 Georgiou, K.; Mittas, N.; Ampatzoglou, A.; Chatzigeorgiou, A.; Angelis, L. What is being Patented in Software Engineering? Empirical Evidence from USPTO. IEEE Softw. **2023**, 1–7. [[Google Scholar](https://scholar.google.com/scholar_lookup?title=What+is+being+Patented+in+Software+Engineering?+Empirical+Evidence+from+USPTO&author=Georgiou,+K.&author=Mittas,+N.&author=Ampatzoglou,+A.&author=Chatzigeorgiou,+A.&author=Angelis,+L.&publication_year=2023&journal=IEEE+Softw.&pages=1%E2%80%937&doi=10.1109/MS.2023.3304349)] [[CrossRef](https://doi.org/10.1109/MS.2023.3304349" \t "_blank)]

 Albino, V.; Ardito, L.; Dangelico, R.M.; Petruzzelli, A.M. Understanding the development trends of low-carbon energy technologies: A patent analysis. Appl. Energy **2014**, 135, 836–854. [[Google Scholar](https://scholar.google.com/scholar_lookup?title=Understanding+the+development+trends+of+low-carbon+energy+technologies:+A+patent+analysis&author=Albino,+V.&author=Ardito,+L.&author=Dangelico,+R.M.&author=Petruzzelli,+A.M.&publication_year=2014&journal=Appl.+Energy&volume=135&pages=836%E2%80%93854&doi=10.1016/j.apenergy.2014.08.012)] [[CrossRef](https://doi.org/10.1016/j.apenergy.2014.08.012" \t "_blank)]

 Trappey, C.V.; Wu, H.-Y.; Taghaboni-Dutta, F.; Trappey, A.J. Using patent data for technology forecasting: China RFID patent analysis. Adv. Eng. Inform. **2011**, 25, 53–64. [[Google Scholar](https://scholar.google.com/scholar_lookup?title=Using+patent+data+for+technology+forecasting:+China+RFID+patent+analysis&author=Trappey,+C.V.&author=Wu,+H.-Y.&author=Taghaboni-Dutta,+F.&author=Trappey,+A.J.&publication_year=2011&journal=Adv.+Eng.+Inform.&volume=25&pages=53%E2%80%9364&doi=10.1016/j.aei.2010.05.007)] [[CrossRef](https://doi.org/10.1016/j.aei.2010.05.007" \t "_blank)]

 Evangelista, A.; Ardito, L.; Boccaccio, A.; Fiorentino, M.; Petruzzelli, A.M.; Uva, A.E. Unveiling the technological trends of augmented reality: A patent analysis. Comput. Ind. **2020**, 118, 103221. [[Google Scholar](https://scholar.google.com/scholar_lookup?title=Unveiling+the+technological+trends+of+augmented+reality:+A+patent+analysis&author=Evangelista,+A.&author=Ardito,+L.&author=Boccaccio,+A.&author=Fiorentino,+M.&author=Petruzzelli,+A.M.&author=Uva,+A.E.&publication_year=2020&journal=Comput.+Ind.&volume=118&pages=103221&doi=10.1016/j.compind.2020.103221)] [[CrossRef](https://doi.org/10.1016/j.compind.2020.103221" \t "_blank)]

 Huang, Z.; Chen, H.; Yip, A.; Ng, G.; Guo, F.; Chen, Z.-K.; Roco, M.C. Longitudinal patent analysis for nanoscale science and engineering: Country, institution and technology field. J. Nanopart. Res. **2003**, 5, 333–363. [[Google Scholar](https://scholar.google.com/scholar_lookup?title=Longitudinal+patent+analysis+for+nanoscale+science+and+engineering:+Country,+institution+and+technology+field&author=Huang,+Z.&author=Chen,+H.&author=Yip,+A.&author=Ng,+G.&author=Guo,+F.&author=Chen,+Z.-K.&author=Roco,+M.C.&publication_year=2003&journal=J.+Nanopart.+Res.&volume=5&pages=333%E2%80%93363&doi=10.1023/A:1025556800994)] [[CrossRef](https://doi.org/10.1023/A:1025556800994" \t "_blank)]

 Sampaio, P.G.V.; González, M.O.A.; de Vasconcelos, R.M.; dos Santos, M.A.T.; de Toledo, J.C.; Pereira, J.P.P. Photovoltaic technologies: Mapping from patent analysis. Renew. Sustain. Energy Rev. **2018**, 93, 215–224. [[Google Scholar](https://scholar.google.com/scholar_lookup?title=Photovoltaic+technologies:+Mapping+from+patent+analysis&author=Sampaio,+P.G.V.&author=Gonz%C3%A1lez,+M.O.A.&author=de+Vasconcelos,+R.M.&author=dos+Santos,+M.A.T.&author=de+Toledo,+J.C.&author=Pereira,+J.P.P.&publication_year=2018&journal=Renew.+Sustain.+Energy+Rev.&volume=93&pages=215%E2%80%93224&doi=10.1016/j.rser.2018.05.033)] [[CrossRef](https://doi.org/10.1016/j.rser.2018.05.033" \t "_blank)]

 Jin, L.; Sun, X.; Ren, H.; Huang, H. Hotspots and trends of biological water treatment based on bibliometric review and patents analysis. J. Environ. Sci. **2022**, 125, 774–785. [[Google Scholar](https://scholar.google.com/scholar_lookup?title=Hotspots+and+trends+of+biological+water+treatment+based+on+bibliometric+review+and+patents+analysis&author=Jin,+L.&author=Sun,+X.&author=Ren,+H.&author=Huang,+H.&publication_year=2022&journal=J.+Environ.+Sci.&volume=125&pages=774%E2%80%93785&doi=10.1016/j.jes.2022.03.037)] [[CrossRef](https://doi.org/10.1016/j.jes.2022.03.037" \t "_blank)]

 Daim, T.U.; Rueda, G.; Martin, H.; Gerdsri, P. Forecasting emerging technologies: Use of bibliometrics and patent analysis. Technol. Forecast. Soc. Chang. **2006**, 73, 981–1012. [[Google Scholar](https://scholar.google.com/scholar_lookup?title=Forecasting+emerging+technologies:+Use+of+bibliometrics+and+patent+analysis&author=Daim,+T.U.&author=Rueda,+G.&author=Martin,+H.&author=Gerdsri,+P.&publication_year=2006&journal=Technol.+Forecast.+Soc.+Chang.&volume=73&pages=981%E2%80%931012&doi=10.1016/j.techfore.2006.04.004)] [[CrossRef](https://doi.org/10.1016/j.techfore.2006.04.004" \t "_blank)]

 Wang, B.; Liu, Y.; Zhou, Y.; Wen, Z. Emerging nanogenerator technology in China: A review and forecast using integrating bibliometrics, patent analysis and technology roadmapping methods. Nano Energy **2018**, 46, 322–330. [[Google Scholar](https://scholar.google.com/scholar_lookup?title=Emerging+nanogenerator+technology+in+China:+A+review+and+forecast+using+integrating+bibliometrics,+patent+analysis+and+technology+roadmapping+methods&author=Wang,+B.&author=Liu,+Y.&author=Zhou,+Y.&author=Wen,+Z.&publication_year=2018&journal=Nano+Energy&volume=46&pages=322%E2%80%93330&doi=10.1016/j.nanoen.2018.02.020)] [[CrossRef](https://doi.org/10.1016/j.nanoen.2018.02.020" \t "_blank)]

 Blei, D.M.; Ng, A.Y.; Jordan, M.I. Latent dirichlet allocation. J. Mach. Learn. Res. **2003**, 3, 993–1022. [[Google Scholar](https://scholar.google.com/scholar_lookup?title=Latent+dirichlet+allocation&author=Blei,+D.M.&author=Ng,+A.Y.&author=Jordan,+M.I.&publication_year=2003&journal=J.+Mach.+Learn.+Res.&volume=3&pages=993%E2%80%931022)]

 Ghaffari, M.; Aliahmadi, A.; Khalkhali, A.; Zakery, A.; Daim, T.U.; Yalcin, H. Topic-based technology mapping using patent data