



Article

PatentInspector: An Open-Source Tool for Applied Patent Analysis and Information Extraction

Konstantinos Petrakis ¹, Konstantinos Georgiou ¹, Nikolaos Mittas ² and Lefteris Angelis ^{1,*}

¹ School of Informatics, Aristotle University of Thessaloniki, 54124 Thessaloniki, Greece; konstpetrakis01@gmail.com (K.P.); konsgeor@csd.auth.gr (K.G.)

² Department of Chemistry, International Hellenic University, 65404 Kavala, Greece; nmittas@chem.ihu.gr

* Correspondence: lef@csd.auth.gr

Featured Application: This work concerns a fully functional and deployed framework for patent analysis. The potential applications of this work range from the exploratory analysis and scoping of patent technologies and themes to the discovery of key companies that invest in specific patent domains. In our study, we present an exploration of patents related to human and project management and demonstrate how the developed tool enables the rapid interpretation of the findings.

Abstract: Patent analysis is a field that concerns the analysis of patent records, for the purpose of extracting insights and trends, and it is widely used in various fields. Despite the abundance of proprietary software employed for this purpose, there is currently a lack of easy-to-use and publicly available software that can offer simple and intuitive visualizations, while advocating for open science and scientific software development. In this study, we attempt to fill this gap by offering PatentInspector, an open-source, public tool that, by leveraging patent data from the United States Trademark and Patent Office, is able to produce descriptive analytics, thematic axes and citation network analysis. The use and interpretability of PatentInspector is illustrated through a use case on human resource management-related patents, highlighting its functionalities. The results indicate that PatentInspector is a practical resource for conducting patent analytics and can be used by individuals with a limited or no background in coding and software development.

Keywords: patent analytics; scientific software development; topic modeling; citation networks



Citation: Petrakis, K.; Georgiou, K.; Mittas, N.; Angelis, L.

PatentInspector: An Open-Source Tool for Applied Patent Analysis and Information Extraction. *Appl. Sci.* **2023**, *13*, 13147. <https://doi.org/10.3390/app132413147>

Academic Editors: Frederico Branco, José Martins and Henrique Mamede

Received: 30 October 2023

Revised: 4 December 2023

Accepted: 8 December 2023

Published: 11 December 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In this era of technological and entrepreneurial progress, an increasing number of companies seek to safeguard their intellectual property. Specifically, the number of annual patent applications has almost tripled in the last two decades, according to a study conducted by the World Intellectual Property Organization (WIPO) [1], rendering patent documents more valuable than ever before. Patents are widely considered as a safe choice for large companies and organizations to secure commercial rights, avoid litigation actions and retain their competitive advantage [2].

The scope and importance of patenting is made clear when considering the large number of patent offices around the world, responsible for receiving, evaluating and granting patent applications. Such offices, with the most prominent ones being the United States Trademark and Patent Office (USPTO), the European Patent Office (EPO) and the China National Intellectual Property Administration (CNIPA), handle the difficult task of processing and analyzing patent documents, examining their objectives and their 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 results [3].

In brief, PA is a field that covers the study of patent documents utilizing proven methodologies and techniques comprising text mining, machine learning and data visualization [4,5]. The results of PA have numerous applications that can be exploited in different

sections within an organization or a business, including R&D management, human resources, mergers and acquisition, company evaluation and competitive intelligence [6]. In addition, PA offers a plethora of opportunities for the extraction of meaningful insights through the application of advanced approaches, such as topic modeling, network analysis and machine learning.

While PA offers valuable insights, it is a time-consuming multi-stage process that requires specific skills to be conducted. Patent documents must be collected from various sources, leveraging APIs offered by the patent offices, if applicable, or by using high-level programming languages and databases. After collecting the documents, they must be preprocessed and filtered to meet certain criteria depending on the research goals and examined domain and, finally, be analyzed using a set of methodologies. While this process may seem simple for a seasoned researcher or an individual with a background in programming, databases and data engineering, there are groups of users, such as industrial actors and business stakeholders, that may not possess these types of skills or knowledge and require PA to be streamlined, automated and free of prior knowledge.

Hence, in recent years, tools that automate the process of PA have emerged and have been utilized within organizations [4], due to the excessive volume of patent documents and the inherent complexity in analyzing them. These tools frequently offer the possibility of identifying and collecting related documents, filtering them based on established criteria and applying PA methodologies. Some of these tools are also offered for advanced scientific purposes and enable researchers from multiple disciplines to overcome the obstacles of PA and easily process patent entries.

However, while PA tools do exist and are in use, to the best of our knowledge, very few of them are available as free, accessible and open-source solutions, with the majority of tools being either proprietary or requiring payment after a short free trial. In addition, the existing open-source PA tools are somewhat complex to navigate, requiring a level of scientific knowledge. Thus, the lack of a flexible, open-source and public PA tool that can cater to the needs of multiple target groups for research purposes is a clear gap in the domain of PA software. Particularly in recent years, and even more so during the COVID-19 pandemic, the programming community has greatly encouraged the principles of open science [7,8] and scientific software development [9,10]. These two concepts combine the need for transparency and openness in all scientific domains along with the creation of accessible software that can process and analyze data using scientific concepts, moving science forward and used primarily for research.

Recognizing (a) the current lack of an intuitive, easy-to-use, public and practical tool for PA, in contrast with the multitude of enterprise solutions, and (b) the growing movement for open science and the development of scientific applications that open additional research avenues to scientists and practitioners that may not be familiar with programming concepts, in this study, we introduce PatentInspector, an extensible open-source tool for PA primarily implemented in Python and deployed publicly for wider use. PatentInspector recognizes the challenges associated with software deployment [11] and leverages containers to reduce them, while providing a collective framework for the retrieval, processing, filtering and analysis of patent records. The tool is designed to be user-friendly, requiring no computer programming knowledge and being accessible by a large range of interested parties. It provides a suite of analytical tools, encompassing descriptive and exploratory alongside topic and citation analysis.

The structure of our study is twofold. First, the PatentInspector public tool is introduced, and its architecture is described. Secondly, a demonstrative analysis is performed on patents utilizing PatentInspector, focusing on administration and management. Specifically, the Cooperative Patent Classification (CPC) group “G06Q10/06” is used for the case study, which encompasses areas such as resource management, workflow optimization, human and project management and enterprise planning and modeling, to demonstrate the capabilities of PatentInspector as a scientific application that performs PA. While there is considerable activity on PA notebooks and applications on software repositories like

Github [12], to the best of our knowledge, this is one of the few PA interface platforms that is easy-to-use and publicly distributed to the scientific and industrial communities.

The rest of the study is organized as follows. In Section 2, background information on PA is offered, focusing on the primary scientific methodologies used in our tool, while also presenting other similar tools. In Section 3, the objectives and key contributions of the study are highlighted. In Section 4, the architecture and development of PatentInspector are presented, while Section 5 serves as a case study of its functionalities. In Section 6, the findings of the case study are discussed, emphasizing the ease of interpretation that the tool provides, and, in Section 7, the main threats to the validity of the study are provided, along with conclusions and suggested future work directions in Section 8.

2. Background Information on Patent Analysis and Tools

The field of PA, generally, concerns the accumulation of patent records from one or multiple patent offices, with the aim of extracting insights and useful information via the application of scientific methodologies, text mining and statistics [5]. The various techniques of PA range from descriptive and exploratory analytics to topic modeling, complex citation networks and machine learning classifiers. In this section, some indicative studies on PA will be presented, focusing on the methodologies supported by PatentInspector, and then the most prominent tools for PA will be analyzed, highlighting their functionalities.

2.1. Patent Analysis Literature

Descriptive/Exploratory Analytics: Several studies have leveraged descriptive statistics to portray the temporal, geographical or technological development of patents in various fields. The results of these studies are either descriptive information about patents (e.g., most prominent organizations) or insights from multivariate methods that explain the relationships between multiple variables. Ardito et al. [13] focus on the IoT domain and explore its trends and dynamics on a country and assignee level, pinpointing the USA and China as prominent countries and Huawei and Qualcomm as the main assignees. Fujii et al. [14] and Tseng and Ting [15] explore the AI domain with knowledge-based methodologies and discover the main technologies and investors in AI trends. In the context of software engineering, Georgiou et al. [16] perform a large-scale analysis on patents from the USPTO to discover the geographical, organizational and technological distributions. Similar analyses have also been conducted in the fields of low-carbon technologies [17], RFID concepts [18], augmented reality [19], nanoscience [20] and photovoltaics [21], indicating that PA as a practice can be efficiently used in multiple application domains and yield practical results. Additional studies have also attempted to combine the use of PA with bibliometrics, enhancing the insights of PA with knowledge derived from the research literature and bibliometric indicators [22–24].

Topic Modeling: Apart from leveraging descriptive statistics and exploratory analysis on patent data, several studies have employed algorithms on patent data that extract topics and thematic axes, pinpointing promising technologies and objectives. Among them, the Latent Dirichlet Allocation (LDA) algorithm, proposed by Blei et al. [25], is by far the most popular when it comes to extracting topics in PA. Due to its efficiency in extracting topics from textual information, LDA has been widely employed in many fields, including vehicular technologies [26,27], where Zhang et al. [27] leveraged a variation of LDA, namely the structural topic modeling (STM) algorithm [28], which has also been employed in [29] for the profiling of hydrogen technologies. Other fields include smart manufacturing [30], sustainable city development [31], data-oriented software [32] and telecommunication patents [33], with the latter reviewing assignee hotspots, based on the extracted topics. Hotspots are particularly important as they emphasize prime investors and technologies and they have also been investigated in a plethora of studies [34–37].

Patent roadmaps, which comprise emerging or trending technologies that pave the road for future patent applications, are also an important part of topic modeling studies. Kim et al. [38] propose a patent development map with a case study in 3D printing,

using LDA, while Ma et al. [39] apply the same process in solar cell technologies. Zhang et al. [40] explore the Blockchain sector to assess technological maturity and forecast trending topics, while a large case study of patents in Australia [41] presents a methodology with semantic information that estimates development for specific topics, with a tailored case study. Finally, Kim et al. [42] leverage CPC clusters in telemedicine patents to evaluate the development of the field.

It should be mentioned that topic modeling has also been employed in studies that explore the profiles of firms, along with their knowledge portfolios [43], and the identification of disruptive technologies that may alter the structure of the market [44], with a case study on photovoltaics.

Citation Networks: Patent citation networks have also been proven to be highly important, based on the related literature, as they portray the interrelations between patent records and uncover the most influential patents or technologies. The most common types of citation networks are the patent-to-patent network, which examines the citations between different patents, and the CPC-to-CPC network, which examines the citations between different patent classes. Patent citation networks have been found to be important indicators in the timely identification of notable patents [45], while their use contributes to the mapping of technological research and discovering deeper connections between different domains [46].

Patent citation analysis has been employed in multiple sectors to find prominent assignees and organizations, technologies and patent entries, including but not limited to vehicle batteries [47], mobile technologies [48], agricultural and natural case studies [49–54], printed electronics [55] and nanotechnology [56]. The diffusion of information in patent citation networks has also been studied [57,58], along with the identification of emerging technologies, their lifecycles [59–61] and the concept of open innovation [62] and whether it is reflected in patent citations.

Technological trajectories are also an aspect that is investigated in patent citations, which can be translated into the forecasting of the evolution of an emerging technology or an established practice based on its status in a citation network. This concept has been studied in patents regarding communication standards and energy devices [63,64], fuel cell research [65] and Blockchain [66]. Finally, several studies focus on assignees along with their associated technologies and their status on patent networks as a sign of competitive advantage, inventive prowess and the largest market share [67,68].

As PA has multiple applications, some studies have also proposed new approaches to exploring patent citations. More specifically, Hu et al. [69] introduce ego citation networks as an alternative means of exploring the citation of patents coupled with bibliographic references. Yang et al. [70] construct a comprehensive patent citation network leveraging direct, indirect, coupling and co-citation metrics, while Chakraborty et al. [71] use exponential random graph models to incorporate social parameters into a patent citation network. Finally, brokerage analysis [72], which exploits triadic relationships, has also been used in patent-to-patent networks [32,57,73].

2.2. Patent Analysis Tools

As mentioned in the Introduction, there are several PA tools that allow the processing of patent records, which are widely used by enterprises and organizations. In Table 1, basic information about the most popular PA tools is presented, highlighting their key characteristics and operations. An inspection of the table reveals that the majority of the tools are, indeed, 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 support 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), which are often

based on AI methodologies, while some of them 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 [74] is an educational suite that leverages data from EPO and focuses on citation networks and clustering. The suite also provides an interface [75] 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 [74], while PatentInspector 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 PatentInspector, 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 that was undertaken before PatentInspector and provided an interface that executed advanced PA with topic modeling.

Table 1. Prominent patent analysis tools.

Tool Name	Sources	Patent Searching	Semantic Analysis	Topic Modeling	Citation Networks	Business Indicators	Descriptive Statistics/ Exploratory Analysis	Public	Proprietary	Free Trial
TopicTracker [77]	-	-	Yes	Yes	No	No	No	No	No	No
TechSpectrogram [78]	PatStat	No	No	Partial	Yes	No	No	No	No	No
PatentMiner [76]	USPTO, JPO, DPMA, IPO, CPD	Yes	Yes	Yes	Yes	No	No	Yes	No	No
Patent2Net [74,75]	EPO	Yes	No	No	Yes	No	Yes	Yes	No	No
PatSeer [79]	Multiple	Yes	No	No	No	Yes	Yes	No	Yes	No
Derwent [80]	Multiple	Yes	Yes	No	-	Yes	Yes	No	Yes	No
Orbit [81]	Multiple	Yes	Yes	No	-	No	Yes	No	Yes	No
IamIP [82]	Multiple	Yes	No	Yes	No	Yes	Yes	No	Yes	No
IPRally [83]	Multiple	Yes	Yes	No	No	Yes	No	No	Yes	Yes
PatBase [84]	Multiple	Yes	Yes	No	Yes	Yes	Yes	No	Yes	On Request
UnifiedPatents [85]	Multiple	Yes	Yes	No	No	Yes	Yes	Yes	No	Yes
SciTech Patent Art [86]	Multiple	Yes	No	Yes	No	Yes	Yes	No	Yes	On Request
Tradespace [87]	Multiple	Yes	Yes	No	No	Yes	No	No	Yes	On Request
AcclaimIP [88]	Multiple	Yes	No	No	Yes	Yes	Yes	No	Yes	Yes
Innography [89]	Multiple	Yes	Yes	No	Yes	Yes	Yes	No	Yes	No
IPLytics [90]	Multiple	Yes	Yes	No	No	Yes	Yes	No	Yes	On Request
Minesoft Origin [91]	Multiple	Yes	Yes	No	No	No	No	No	Yes	On Request
Octimine [92]	Multiple	Yes	Yes	No	No	Yes	Yes	No	Yes	On Request
Patent Inspiration [93]	Multiple	Yes	Yes	No	Yes	Yes	Yes	No	Yes	On Request
PatentSight [94]	Multiple	Yes	Yes	Yes	No	Yes	Yes	No	Yes	On Request
PatSnap [95]	Multiple	Yes	Yes	No	No	Yes	Yes	No	Yes	Yes
PatentInsight [96]	Multiple	Yes	Yes	Partial	Partial	Yes	Yes	No	Yes	On Request

Table 1. Cont.

Tool Name	Sources	Patent Searching	Semantic Analysis	Topic Modeling	Citation Networks	Business Indicators	Descriptive Statistics/ Exploratory Analysis	Public	Proprietary	Free Trial
PQAI [97]	Multiple	Yes	Yes	No	No	No	No	No	No	Yes
PatZilla [98]	EPO (mainly)	Yes	No	No	No	No	No	Yes	No	No
Google Patents [99]	Multiple	Yes	No	Partial	No	No	Yes	Yes	No	No
FreePatents [100]	USPTO, EPO, JPO, WIPO	Yes	No	No	No	No	No	Yes	No	No
Relucura (TechTraker, TechExplorer, Enterprise Web tool) [101]	Multiple	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	No
The Lens [102]	Multiple	Yes	No	No	Partial	No	Yes	Yes	Yes	No
PatentR [103]	Multiple	No	No	No	No	No	Yes	Yes	No	Yes
Sumobrain [104]	Multiple	Yes	No	No	No	No	No	Yes	No	Yes
PatentAnalyzer [105]	Multiple	Yes	Yes	No	Yes	No	No	No	Yes	On Request
Patexia PatentAnalyzer [106]	Multiple	Yes	Yes	No	No	Yes	Yes	No	Yes	On Request
PatentInspector	USPTO	Yes ¹	No	Yes	Yes	No	Yes	Yes	No	Yes

¹ Note that in PatentInspector, the data source (USPTO) is downloaded in the database and not retrieved in real time.

The remaining free PA tools (PatZilla, FreePatents and GooglePatents) are not PA tools in the typical sense, as they mainly provide advanced search engines for the retrieval of patent documents. Thus, their PA capabilities are minimal and they cannot be considered similar to PatentInspector, which employs established scientific concepts and targets all types of users. GooglePatents [99] 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).

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. PatentInspector emerges to cover this deficit, with results from the USPTO while also offering different methodologies, efficient visualizations and interpretable insights. In addition, PatentInspector 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.

3. Objectives and Contribution

The main objective of this study was to create PatentInspector, a user-friendly tool designed for both scientific research and everyday use. Our goal is to provide a resource that is open to any individual and simplifies the process of PA, offering scientific concepts in an easily digestible manner. The novelty of the tool that has been developed is that, in contrast to the plethora of proprietary software, it focuses on research aspects and semantic insights from patents by leveraging topic modeling and citation analysis methodologies and visualizations. Thus, it offers a new perspective on patent activity, and it can be utilized by mainstream users in combination with insights from other PA solutions.

Overall, the primary contributions of PatentInspector are the following.

C1. Provide accessible PA: The goal behind PatentInspector is to widen the reach of PA, making it accessible to a wider audience without the need for expertise in legal frameworks, computer programming or data science. It is our belief that, as in many domains, the wider public is unable to extract insights and analyze patent data due to existing software being primarily proprietary and data retrieval pathways requiring coding knowledge. PatentInspector strives to provide a solution to these problems, offering a solution that automates data retrieval and guides the potential user to the analyses that it performs.

C2. Bridge the gap between PA complexity and knowledge: Complementary to C1, PatentInspector seeks to minimize the inherent complexity of the PA field and enable individuals with a limited or no programming background to be able to use, even in an elementary fashion, a tool that can process patent records and extract results. The developed solution, while offering some additional functionalities for more experienced users, requires no advanced knowledge of PA, topic modeling or statistics, thus allowing anyone to use it effectively. We aspire for PatentInspector to become a valuable resource, enabling numerous individuals to gain insights into their areas of interest within the patent landscape. Based on its design, the developed platform can be applicable across various domains and accessible to individuals from diverse backgrounds.

C3. Flexible, open-source tool for PA: As mentioned in Section 2, a large number of existing PA software programs are proprietary and must be purchased or subscribed to. This, in turn, limits the pool of users that can utilize them, while the learning curve may be high. Hence, with respect to the rise of scientific software development and open science, we offer an extensible, public tool for PA that not only is publicly available regarding the usage and modification of the source code but is also flexible in its design and primary functions.

C4. Favor simplicity, encourage engagement: At its core, the proposed tool was designed to be simple and easy-to-use. The frontend component is composed of visualizations that do not contain complex information, while more sophisticated concepts are not forced

on the regular user but can be leveraged by more experienced users. Thus, PatentInspector has the potential to achieve high engagement by any user due to its simplistic yet sophisticated nature.

The proposed tool has practical application value for various interested parties, who can use it for different objectives and purposes. The different implications and target groups are presented below. In addition, Figure 1 indicates the different ways that PatentInspector can be operated by various individuals in a concise manner. However, it is crucial to note that while PatentInspector provides insights into the patent landscape, it should not be the sole basis for important decisions. It also does not aim to replace manual PA, lacking certain features and the expertise of researchers. Finally, it must be noted that the current version of the tool only uses USPTO as its data source, thus limiting the results towards the US. In future versions of the tool, we plan to include more data sources.

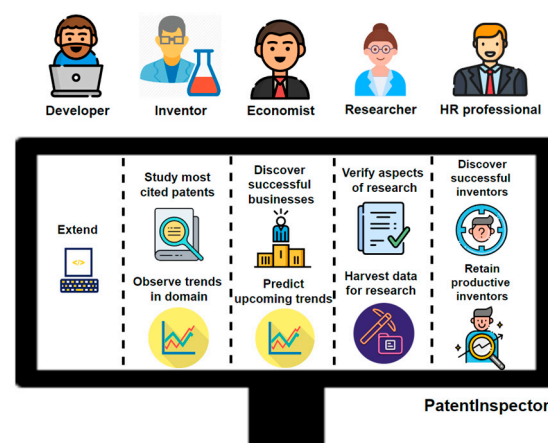


Figure 1. PatentInspector use cases from various individuals.

11. Developers: PatentInspector follows established architecture schemas, and it is fully open-source. Developers and programmers can utilize the source code, enrich and extend it with additional features and capabilities and develop the application. The code has been structured so as to encourage novice programmers to enhance their software development skills but also experienced professionals to modify it according to their preferences.

12. Patent inventors: The developed tool can certainly benefit individuals that have accomplished an invention and wish to patent it. Primarily, it serves as a practical way to identify frequently cited patents in their research field, offering valuable insights and trends while also revealing whether their invention is innovative and can potentially be granted. It should be emphasized, though, that, currently, PatentInspector only supports patent grants from the USPTO, so any results would inevitably be skewed towards the US.

13. Economists: Individuals that deal with the stock market, return-of-investment and economic deals can leverage PatentInspector to observe patent trends, focusing on specific organizations, scoping successful businesses and emerging patent fields to predict upcoming trends and make informed decisions.

14. HR departments and policymakers: PatentInspector has also an important societal aspect in its functionalities, as it can be used in conjunction with other tools and software for business intelligence and skills analysis, to discover successful inventors. This in turn can lead HR departments and policymakers within organizations to extract insights for talent acquisition, by scouting active inventors and recruiting them or retaining active personnel in their own organizations.

15. Researchers: Researchers are another important group that can use PatentInspector, as PA is a highly active field in research, with valuable insights [3,6]. Hence, researchers with a grasp of scientific methodologies can not only use PatentInspector as a validity check, when conducting manual PA, but they can also employ its various functionalities to accelerate their research and leverage the results for more complex algorithms. In

addition, PatentInspector is an excellent alternative for harvesting patent records from a selected domain, with a variety of features. However, we should once again point out that PatentInspector cannot replace global patent databases, as, in its latest release, it only retrieves data from the USPTO.

4. Architecture and Workflow

PatentInspector has the structure of a standard web application, consisting of a frontend and a backend component, each with a distinct role. The backend component is developed in Python 3.11, using the Django framework [107], and is responsible for storing data, conducting computations and managing the application's core functionality. The efficient handling of the aforementioned operations is made possible by utilizing the Postgres relational database, which stores and retrieves data through complex SQL queries generated by Django's Object Relational Mapper (ORM). Additionally, the backend component provides management commands and performs necessary preprocessing procedures to streamline the tool administration. On the other hand, the frontend component, built using the JavaScript Vue framework [108], plays a crucial role in presenting the data to users in an intuitive and interactive manner. It is the part of the application that users directly interact with, providing an interface for accessing the information processed by the backend. In the frontend component, users can interact with the interface and generate PA reports while also being able to access previous reports that they have created. The overall architecture of the tool is presented in Figure 2.

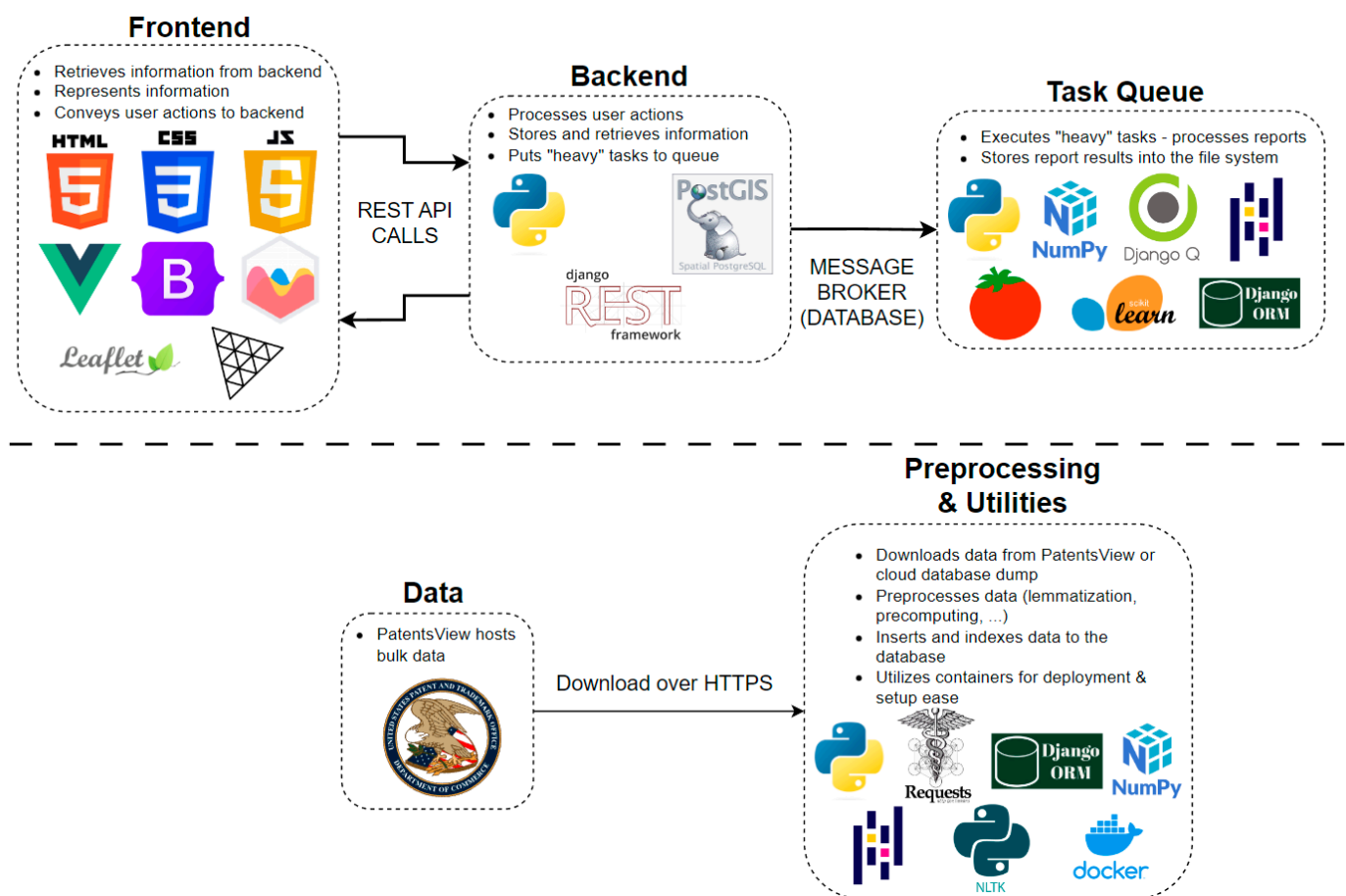


Figure 2. PatentInspector architecture.

The Python programming language was chosen for its widespread popularity, especially in the realm of scientific computing [109], as well as its flexible functionality and maintainability.

4.1. Data Collection, Preprocessing and Storage

PatentInspector operates on patent record data offered free of charge by the USPTO. More specifically, the database used in PatentInspector relies on bulk data available in the PatentsView platform of USPTO [110], which serves as a repository of all USPTO-registered and granted patents and is updated regularly. The tool includes a management utility named “USPTO”, showcased in Figure 3, that automates the process of downloading, decompressing, preprocessing and inserting the data into the database. This repository is organized in tables, with each table containing a different aspect of patent records (e.g., patent classes, patent inventors, etc.)

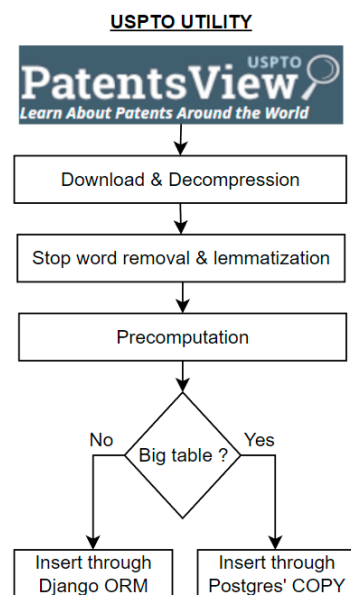


Figure 3. An overview of the USPTO management utility.

PatentInspector emphasizes only targeted tables made available from USPTO and, more specifically, only those regarding granted patents, while it does not retrieve those who have applied to USPTO but have not yet received a patent grant. This was a conscious decision based on the rationale that applied patents may be rejected by the patent office and would thus hold reduced importance in the collected data [111]. The different fields and tables retrieved are presented in Table 2.

Table 2. Fields retrieved from the USPTO.

Table Name	Description	Fields
g_location_disambiguated	Locations of inventors and assignees	location_id, country, state, city, longitude, latitude, county_fips, state_fips
g_cpc_title	CPC schema	cpc_subclass, cpc_subclass_title, cpc_group, cpc_group_title, cpc_class, cpc_class_title
g_patent	Patent information for granted patents	patent_id, patent_type, patent_date, patent_title, patent_abstract, num_claims, withdrawn
g_application	Application information of granted patents	patent_id, filing_date
g_figures	Number of drawings and drawing sheets of granted patents	patent_id, num_figures, num_sheets
g_cpc_current	CPC classification of granted patents	patent_id, cpc_group
g_pct_data	PCT data of granted patents	patent_id, published_or_filed_date, filed_country, pct_doc_number, pct_doc_type

Table 2. *Cont.*

Table Name	Description	Fields
g_inventor_disambiguated	Inventors of granted patents	patent_id, inventor_id, location_id, disambig_inventor_name_first, disambig_inventor_name_last
g_assignee_disambiguated	Assignees of granted patents	patent_id, assignee_id, location_id, disambig_assignee_name_first, disambig_assignee_name_last, disambig_assignee_organization
g_us_patent_citation	US patent citations of granted patents	patent_id, citation_patent_id, citation_date
g_foreign_citation	Foreign patent citations of granted patents	patent_id, citation_application_id, citation_date, citation_country
g_ipc_at_issue	IPC codes at the time of the patent grant ¹	patent_id, ipc_group

¹ The USPTO does not keep track of current IPC codes but the ones at the time of the grant.

After the tables of interest, containing information about roughly eight million patents, are downloaded, an automated preprocessing procedure takes place. The preprocessing deployed in PatentInspector involves stop word removal and the lemmatization of text fields such as the patent's title or abstract, to facilitate and accelerate the text analysis performed in later 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 stored cache. Table 3 summarizes all the precomputed fields that aid the throughput of the application.

Table 3. Precomputed database fields of PatentInspector.

Field	Type	Table
Granted year	Integer	Patent
Application year	Integer	Patent
Years to get granted	Float	Patent
Title word count without processing	Integer	Patent
Title word count with processing	Integer	Patent
Abstract word count without processing	Integer	Patent
CPC groups count	Integer	Patent
Assignee count	Integer	Patent
Inventor count	Integer	Patent
Incoming citation count	Integer	Patent
Outgoing citation count	Integer	Patent
Citation year	Integer	Patent Citation
Representation	Varchar	PCT Data
IPC groups counts	Integer	Patent

After the preprocessing is finished, the entire set of patent records is inserted into the database. This is achieved through two different approaches depending on the size of the data. For small tables such as Location, the Django ORM is leveraged to insert the data. For larger tables such as Patent, the preprocessed chunks are stored in a bulk CSV in the file system, which is later loaded into the database using Postgres' COPY command, resulting in a significant performance boost. The schema of the database for computational-related and user-related tables is shown in Figures 4 and 5, respectively.

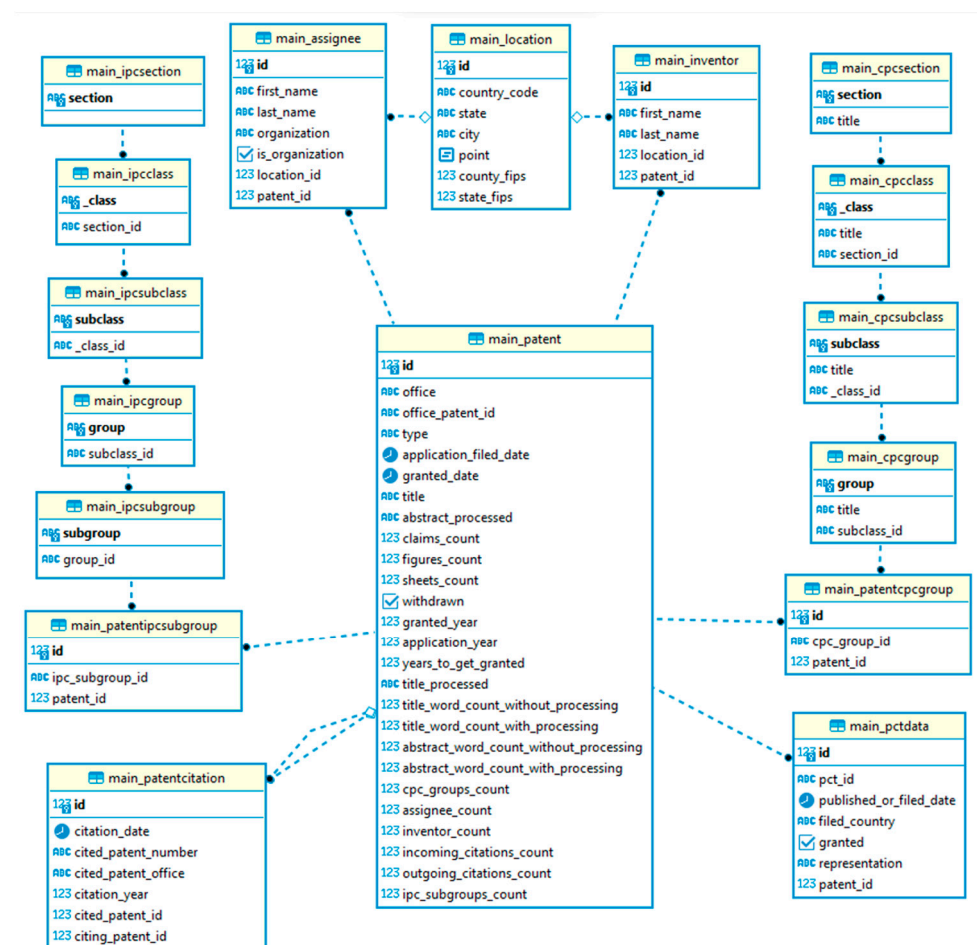


Figure 4. The ER diagram of the computational-related tables.

It is important to highlight that more experienced individuals that wish to deploy PatentInspector on their local machines, rather than running the publicly deployed version, need not depend on the USPTO utility. PatentInspector provides an alternative tool known as the “Load Database” utility that facilitates the retrieval of a highly compressed and indexed dump from the cloud, subsequently loading it into the database. This process typically results in a significant reduction in waiting time, from approximately ten hours to just one hour, on a standard personal computer with a conventional network connection, while this utility is also employed in configuring containers for deployment purposes.

It should also be noted that expanding PatentInspector to incorporate data from other patent offices would involve developing a counterpart utility to USPTO that would correspond to each of the targeted patent offices. This utility would handle the downloading, preprocessing and data insertion tasks.

PatentInspector is specifically designed to support multiple users simultaneously. Consequently, the storage of user-specific data is imperative; hence, user authorization and password reset functionalities are integral features. Users have the ability to generate reports that contain PA insights and access only their individual reports. These reports encompass criteria for patent filtering and include metadata pertinent to the analysis, such as creation dates. The analysis results are stored in the file system, utilizing a combination of JSON and Excel file formats. Further information on PatentInspector reports can be found in the “Computation” subsection (Section 4.2).

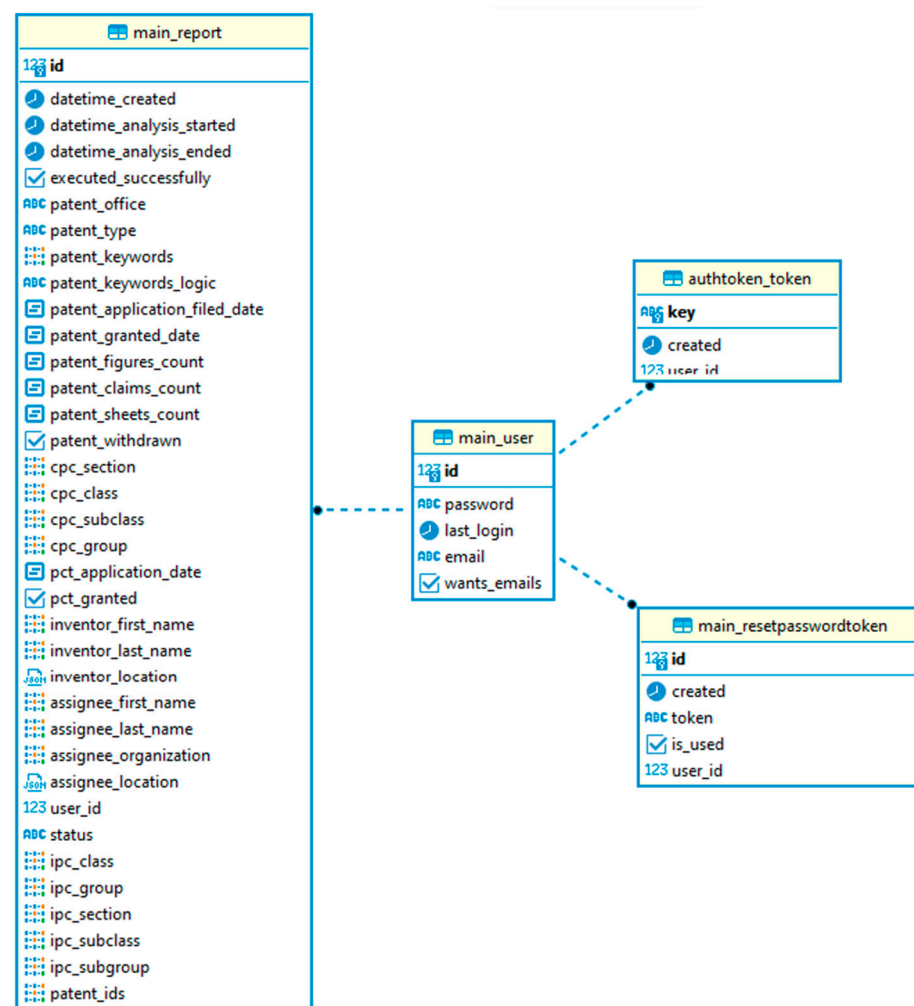


Figure 5. The ER diagram of the user-related tables.

4.2. Computation

The report entity serves as the central element in the user experience, driving computations. When a user initiates a new report or interacts with its results, triggering additional computations, they effectively add a new task to the task queue of PatentInspector. The use of a task queue, rather than executing computations immediately upon request, is essential because analyses can consume up to twenty minutes on an average computer. The task queue of PatentInspector periodically polls for new tasks and executes them in the background when it has available resources. To relieve of the user of having to wait for his/her report to be completed, once tasks are terminated, users are informed via email, based on their preference, that their analysis has been finished.

In PatentInspector, the term “task” refers to functions and their corresponding arguments that are executed at a deferred point in time. These tasks primarily consist of functions integrated with Django ORM code, which ultimately generate complex SQL queries sent to the database. In certain instances, tasks may include code from computational libraries to handle computations that cannot be carried out within the database system, such as topic analysis.

PatentInspector currently implements two tasks, namely “process report” and “topic analysis”. In the “process report” task, all computations are executed using default parameters. For instance, the default setting for the number of topics in topic analysis is ten topics. While users can effectively use the tool with the default parameters, the tool allows users to modify their queries (e.g., change the number of topics) and produce alternative results.

The results of these computations are saved in two files: a JSON file containing all computational outputs and an Excel file containing the patents and their associated information. Users can easily download the Excel file for further manual analysis and tasks. The “topic analysis” task essentially executes the topic analysis methodologies utilizing user-provided parameters, effectively replacing the existing JSON result file. Programmatically extending PatentInspector to provide additional reactivity in the report results is straightforward, with supplementary sub-tasks like “patent analysis” having to be developed.

4.3. API and Interconnection

PatentInspector employs a REST API, which is accessed by the frontend application for user and task management, as well as data retrieval. The tool provides Swagger documentation, which is automatically accessible from the local server when running in a development environment at the “/swagger” URI. The API endpoints are summarized in Table 4.

Table 4. The endpoints of the PatentInspector API.

Endpoint	Method	Short Description
/assignees	GET	Paginated retrieval of assignee names by query params ¹
/cpc/sections	GET	Retrieval of all CPC sections ¹
/cpc/classes	GET	Retrieval of all CPC classes ¹
/cpc/subclasses	GET	Paginated retrieval of CPC subclasses by query params ¹
/cpc/groups	GET	Paginated retrieval of CPC groups by query params ¹
/ipc/section	GET	Retrieval of all IPC sections ¹
/ipc/classes	GET	Retrieval of all IPC classes ¹
/ipc/subclasses	GET	Paginated retrieval of IPC subclasses by query params ¹
/ipc/groups	GET	Paginated retrieval of IPC groups by query params ¹
/ipc/subgroups	GET	Paginated retrieval of IPC subgroups by query params ¹
/inventors	GET	Paginated retrieval of inventor names by query params ¹
/report	GET	Paginated retrieval of report metadata of the authenticated user
/report	POST	Report creation with given criteria in the request body.
/report/:id	GET	Retrieval of report information
/report/:id	DELETE	Deletion of report
/report/:id/download_patents_excel	GET	Streams the Excel file containing the exported patents of a report
/report/:id/get_patents	GET	Paginated retrieval of patents analyzed in the report
/report/:id/topic_analysis	POST	Adds a new topic analysis task to the task queue
/user	POST	Creates a new user based on the information in the request body
/user/ask_reset_password	POST	Sends an OTP in the email of request body if it corresponds to a user
/user/get_data	GET	Returns information for the authenticated user
/user/login	POST	Authenticates the user by returning a token if valid credentials were given in the request body
/user/reset_password	POST	Sets the password to a new one if a valid OTP was given in the request body
/user/update_email	POST	Updates the email of the authenticated user to the one set in the request body
/user/update_password	POST	Updates the password of the authenticated user if the old password in request body is valid
/user/update_wants_emails	POST	Updates the status of receiving emails for the authenticated user based on the request body

¹ Used to search and show valid options for user to select when creating the report in a dropdown menu.

4.4. Features List and Users’ Perspective

User management and verification is an integral part of PatentInspector. The tool allows users to be registered, authenticated and update their credentials and preferences. In Figure 6, we provide an overview of the user management-related windows of the application, indicating that users can change passwords, log in and register while also being able to notify the tool to alert them via email when the analysis is completed, thus eliminating the need of leaving the application open to complete its computations. PatentInspector allows users to create reports based on a comprehensive set of filters targeting various aspects of

the patent ecosystem. The basic idea behind the tool is that users can select which patent records to be analyzed using multiple criteria, ranging from the grant year or keywords in the patent title/abstract to inventor locations or names. Table 5 provides a concise overview of the available filters within the report construction form from a programmer's perspective, while Figure 7 provides the same information from a user's perspective. After users choose their PA criteria and submit the form, they are taken to their report list, presented in Figure 8. There, they can easily check report metadata, delve into previous report results or delete reports as needed.

Figure 6. User management views in PatentInspector.

Figure 7. The filters of the report construction form in PatentInspector from a user's perspective.

Table 5. The filters of the report construction form in PatentInspector from a programmer’s perspective.

Filter	Type	Entity	Example
Office	String	Patent	"US"
Type	String	Patent	"utility"
Keywords	Array<String>	Patent	["IoT", "Big data"]
Keywords logic	String	Patent	"&"
Application filed date	Object	Patent	{"lower": "1912-06-23", "upper": null}
Granted date	Object	Patent	{"lower": "1912-06-23", "upper": null}
Figures count	Object	Patent	{"lower": 2, "upper": 5}
Claims count	Object	Patent	{"lower": 2, "upper": 5}
Sheets count	Object	Patent	{"lower": 2, "upper": 5}
Withdrawn	Boolean	Patent	false
Sections	Array<String>	CPC	["A", "B"]
Classes	Array<String>	CPC	["A01", "A21"]
Subclasses	Array<String>	CPC	["A01B", "A01C"]
Groups	Array<String>	CPC	["A01B1/022", "A01B1/024"]
Sections	Array<String>	IPC	["A", "B"]
Classes	Array<String>	IPC	["A01", "A21"]
Subclasses	Array<String>	IPC	["A01B", "A01C"]
Groups	Array<String>	IPC	["A01B 1", "A01C 1"]
Subgroups	Array<String>	IPC	["A01B 1/022", "A01B 1/024"]
Application filed date	Object	PCT	{"lower": "1912-06-23", "upper": "2023-06-10"}
Granted	Boolean	PCT	true
First name	Array<String>	Inventor	["Alan", "Ada"]
Last name	Array<String>	Inventor	["Turing", "Lovelace"]
Location	Object	Inventor	{"lat": 40.633, "lng": 22.956, "radius": 500}
First name	Array<String>	Assignee	["Alan", "Ada"]
Last name	Array<String>	Assignee	["Turing", "Lovelace"]
Organization	Array<String>	Assignee	["Example Corporation"]
Location	Object	Assignee	{"lat": 40.633, "lng": 22.956, "radius": 500}

PatentInspector

All Reports

Showing 4 report(s) in total.

Inspect	Created	Analysis started	Analysis ended	Analysis duration	Delete
<button>Inspect</button>	10/2/2023, 1:26:11 PM	10/2/2023, 1:26:12 PM	10/2/2023, 1:33:25 PM	00:07:13	<button>Delete</button>
<button>Inspect</button>	9/30/2023, 6:29:48 PM	9/30/2023, 6:29:49 PM	9/30/2023, 6:31:09 PM	00:01:19	<button>Delete</button>
<button>Inspect</button>	9/30/2023, 4:03:59 PM	9/30/2023, 4:04:00 PM	9/30/2023, 4:04:16 PM	00:00:16	<button>Delete</button>
<button>Inspect</button>	9/24/2023, 5:23:13 PM	9/24/2023, 5:23:14 PM	9/24/2023, 5:24:10 PM	00:00:55	<button>Delete</button>

Previous **1** Next Showing 4 results in 1 pages

Figure 8. The report list user interface of PatentInspector.

4.5. Analysis Tabs

The analysis conducted by PatentInspector is organized into three primary tabs, namely the Descriptive Analysis Tab, the Thematic Analysis Tab and the Network Analysis Tab. In this section, we present each tab from the perspective of the user, analyzing the functionalities that they provide.

The Descriptive Analysis Tab of PatentInspector is organized into three distinct sections, each serving a specific purpose. Firstly, the “Basic Statistical Measures” section offers a table featuring statistical measures for a range of variables. Secondly, the “Variables

Over Time” section provides insights through various time series representations. Lastly, the “Information for Each Entity” section presents data distributions tailored to different aspects of PA, ensuring an inclusive view of the data for individual entities. In Figure 9, the Descriptive Analysis Tab is presented, while Table 6 summarizes the derived statistics.

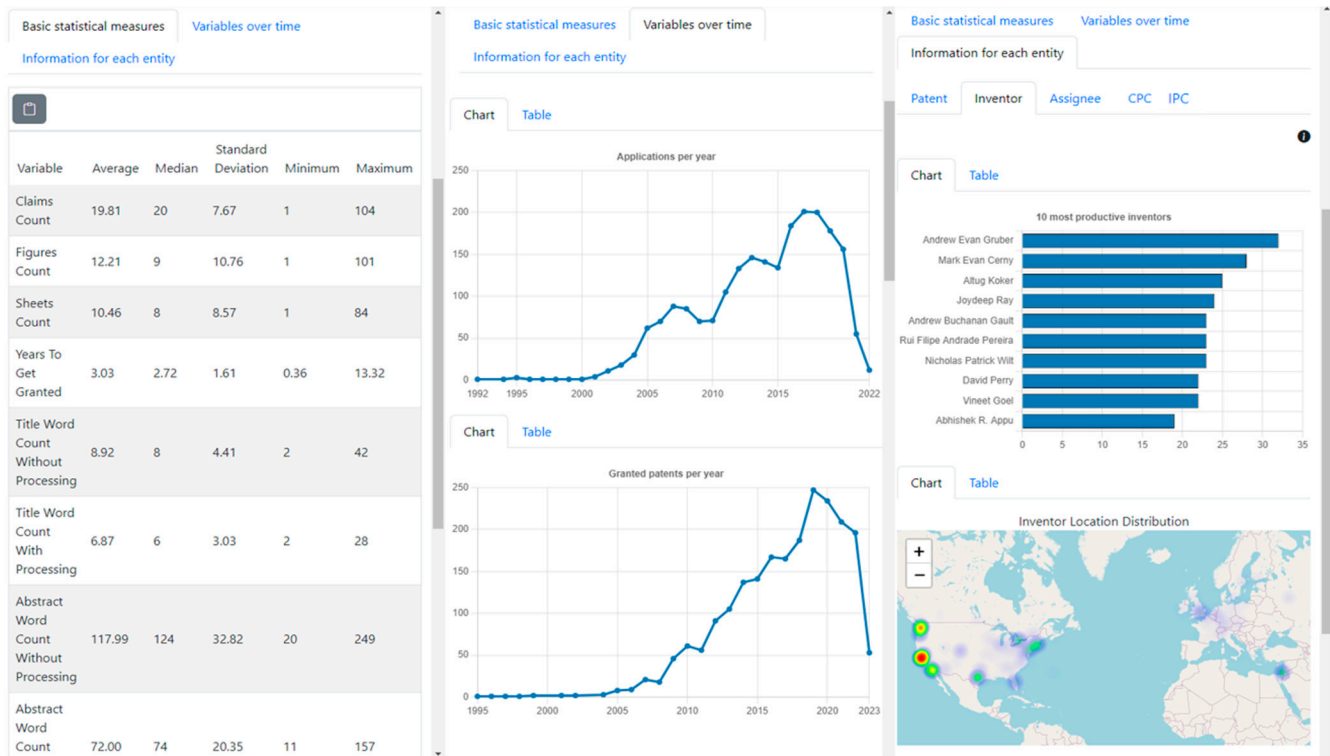


Figure 9. The user interface of the Descriptive Analysis Tab of PatentInspector.

The Topic Analysis Tab consists of three main components. First, there is a form that allows users to adjust the criteria for topic analysis. These criteria include the choice of topic analysis method (the tool currently supports the Latent Dirichlet Allocation (LDA) method and the Nonnegative Matrix Factorization (NMF) method [112]), the number of topics, the words per topic, the date range for analysis and parameters like the removal of the most common words (for LDA) or the maximum document frequency (for NMF). The second component displays a scatter plot and its corresponding table, categorizing topics as “emerging,” “dominant,” “declining” or “saturated” based on the methodology outlined in [36], which relies on the patent share (the number of patents in each topic) and the Compound Annual Growth Rate (CAGR) of the patent share. Lastly, the third component presents topic details, including word weights and the number of patents in each topic. Figure 10 presents a detailed overview of the Topic Analysis Tab.

Finally, the tool includes a Network Analysis Tab that showcases the most cited patents across local and global networks. Furthermore, it provides an interactive 3D graph representation for the local citation network. These elements are presented in Figure 11.

Another important point is that within the framework of PatentInspector, a Patents Tab is featured (presented in Figure 12). This tab hosts a sortable table containing the patents that have been retrieved with the filters applied by the user, accompanied by the functionality to download a comprehensive Excel file. This file encompasses most of the pertinent information for the patents that have undergone filtration based on the specified criteria submitted when creating the report. The Patents Tab is highly important as it enables users with a richer background and knowledge to conduct a more in-depth analysis of the patents on their own terms. However, this does not limit users with limited knowledge from experimenting with the tool or downloading the extracted patents for additional analysis.

Table 6. Features of the Descriptive Analysis Tab.

Section	Content
Basic statical measures	<p>Table of statistical measures (avg., med., std., min, max) for the following variables:</p> <ul style="list-style-type: none"> • Claims count • Figures count • Sheets count • Assignee count • Inventor count • Incoming citation count • Outgoing citation count • CPC groups count • IPC subgroup count • Wait time for grant in years • Title word count ¹ • Abstract word count ¹
Variables over time	<p>Line charts that depict the evolution of the following variables over time:</p> <ul style="list-style-type: none"> • Application count • Grant count • Grants per type ² • Grants per office ² • PCT grants count • Grants per CPC section ² • Outgoing citation count • Incoming citation count
Information for each entity	<p>Distributions and observations for the following entities:</p> <ul style="list-style-type: none"> • Patent <ul style="list-style-type: none"> o PCT application status o Patent type o Patent office • Inventor <ul style="list-style-type: none"> o 10 most prolific inventors o Inventor locations ³ • Assignee <ul style="list-style-type: none"> o 10 most patent-holding assignees o Assignee type o Assignee location ³ • CPC Patent distribution across: <ul style="list-style-type: none"> o CPC sections o 5 most popular CPC classes o 5 most popular CPC subclasses o 5 most popular CPC groups • IPC Patent distribution across: <ul style="list-style-type: none"> o IPC sections o 5 most popular IPC classes o 5 most popular IPC subclasses o 5 most popular IPC groups o 5 most popular IPC subgroups

¹ Word counts both with and without stop word removal are included. ² Multiple lines are shown within the same line chart. ³ Shown via global heatmaps.

Extracted **10** topics from the corpus using the **LDA** model. Showing **10** words per topic.

The dates used for the CAGR calculation utilized in the topic classification are the following: **3/30/2015–3/28/2020**.

The **20** most common words were removed.

Coherence score: **0.513**

You can conduct a new topic analysis within the same report by filling out the form below:

Topic Analysis Form

Method: **LDA**

Number of topics: **10**

Number of words per topic: **10**

Dates for classification: **03/30/2015** – **03/28/2020**

Remove most common words: **20**

Submit

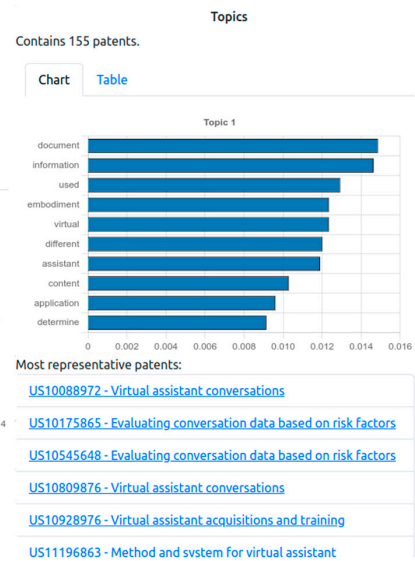
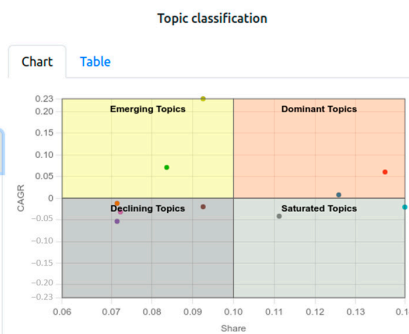


Figure 10. The user interface of the Topic Analysis Tab of PatentInspector. The colored dots correspond to the detected topics and each topic has been assigned a quadrant based on its Patent Share and CAGR values.



Figure 11. The user interface of the Network Analysis Tab of PatentInspector. Red dots correspond to patents and grey lines connect patents that cite or are cited by other patents.

Download Excel The excel file contains more information than the table below.

Id	Office	Office Patent Id	Type	Application Filed Date	Granted Date	Title	Abstract Processed	Claims Count	Figures Count	Sheets Count
2036	US	10002021	utility	2012-07-20	2018-06-19	Deferred preemption techniques for scheduling graphics processing	disclosure directed deferred preemption technique scheduling graphic processing unit gpu	62	11	9

Figure 12. The user interface of the Patents Tab of PatentInspector.

5. Case Study and Validation

To effectively showcase its functionalities and usefulness, in this section, PatentInspector is employed to perform a PA focused on the CPC group “G06Q10/06”. This group spans a wide array of domains, including resource management, workflow optimization and human and project management, as well as enterprise planning and modeling. To validate the findings, a comparison of the results is made with a replication of this case study using the Lens software [102], which is hailed as an established source both for patent data retrieval and for PA insights [113–116]. The results indicate that the descriptive insights and citation analysis extracted by PatentInspector largely correspond with the results from Lens, indicating that the constructed tool produces valid PA outputs. However, it should be taken into account that all comparisons were made with patents from the USPTO and not from the global patent landscape. The detailed description and comparison of the tool with Lens [102] can be found in the Supplementary Materials, along with three additional case studies conducted to further validate PatentInspector.

Our analysis, consisting of 13,424 patents retrieved from the filtering system using the “CPC = G06Q10/06” filter using exact matching, commences with a descriptive analysis, provided by the Descriptive Analysis Tab, starting with statistical measures, followed by an exploration of the variables over time, an investigation of entity-specific data and a subsequent topic analysis from the Topic Analysis Tab, concluding with a citation analysis from the Network Analysis Tab. In Table 7, we present basic statistical measures.

Table 7. Statistical measures for G06Q10/06.

Variable	Avg	Med	Std	Min	Max
Claims count	20.05	19	14.22	1	539
Figures count	16.1	10	29.01	0	2045
Sheets count	14.25	9	29.61	0	2037
Years to get granted	3.91	3.48	2.19	0.23	19.23
Title word count without processing	8.93	8	4.40	1	52
Title word count with processing	6.65	6	2.9	1	32
Abstract word count without processing	116.58	119	39.33	8	344
Abstract word count with processing	70.29	71	23.96	0	222
CPC groups count	7.30	5	13.95	1	205
IPC Subgroups count	3.37	2	4.97	1	104
Assignee count	1.01	1	0.14	1	6
Inventor count	3.17	3	2.33	1	63
Incoming citation count	25.24	6	66.42	1	2091
Outgoing citation count	26.09	9	69.3	1	2328

In this context, it is important to highlight that the distributions of applications and grants tend to fall around 2021–2023 due to the absence of data from PatentInspector. As previously detailed in Section 4.1, PatentInspector exclusively handles granted patents. It is noteworthy that the statistical table indicates an average pre-grant duration of 3.91 years.

Consequently, it is reasonable to infer that patents submitted within the last three years are likely not included in the PatentInspector database. Additionally, USPTO is the only patent office that granted patents for G06Q10/06 solely because PatentInspector contains only patents from USPTO currently.

Upon an examination of the charts (Figures 13–20), it becomes evident that the domains associated with HRM present an upwards trajectory as the patent grants and applications are consistently on the rise. Regarding PCT status, while it presents an upwards trend based on Figure 17, 12,517 patents (93.2%) have not applied for PCT, while only 907 (6.8%) of the total patents have been granted PCT status.

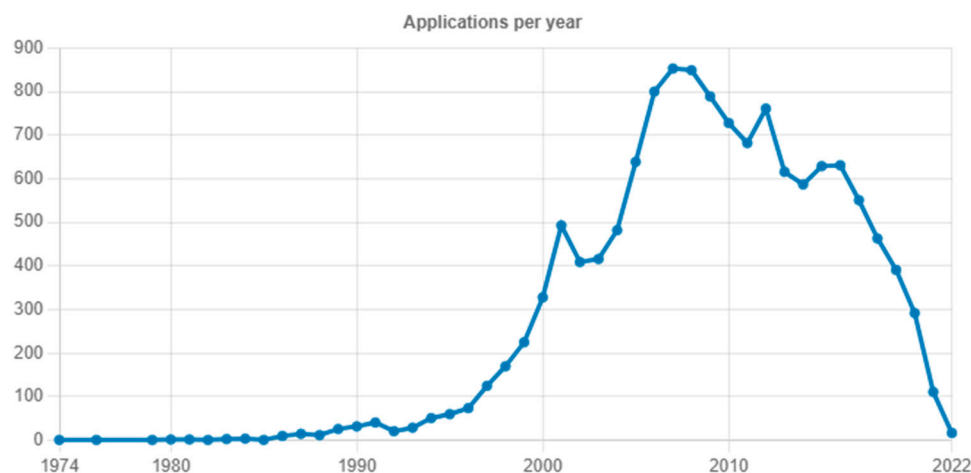


Figure 13. Applications per year for G06Q10/06.

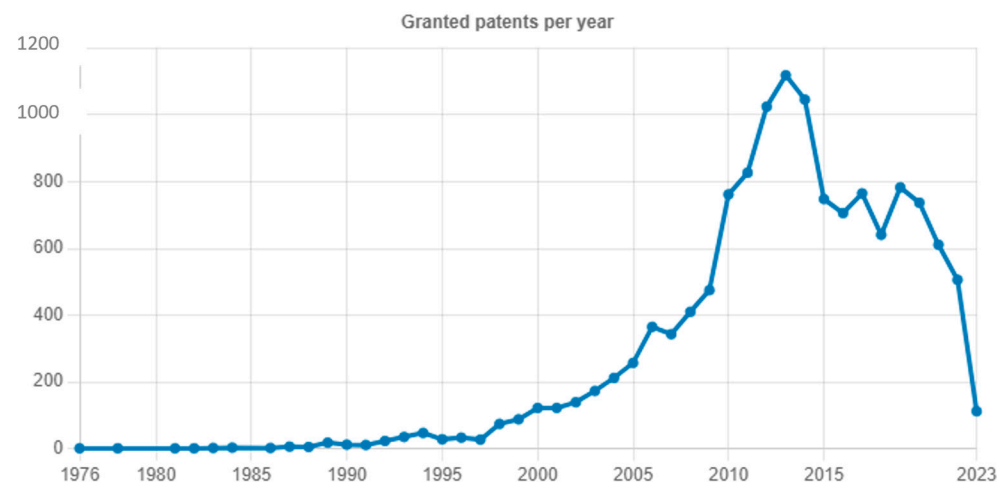


Figure 14. Granted patents per year for G06Q10/06.

The most prolific inventors (Table 8) typically possess a minimum of 27 patents each within the domain of human and resource management. The majority of these inventors are situated in the United States, Japan, the United Kingdom, Ireland, Germany and Israel, as presented in Figure 21. It is important to note that there may be a notable bias in these statistics due to the database of PatentInspector being limited to patents granted by the US patent office, thus skewing the results towards the US. However, the presence of diverse inventors from different countries indicates that HRM is indeed a globally studied field, with multiple individuals interested in patenting their inventions.

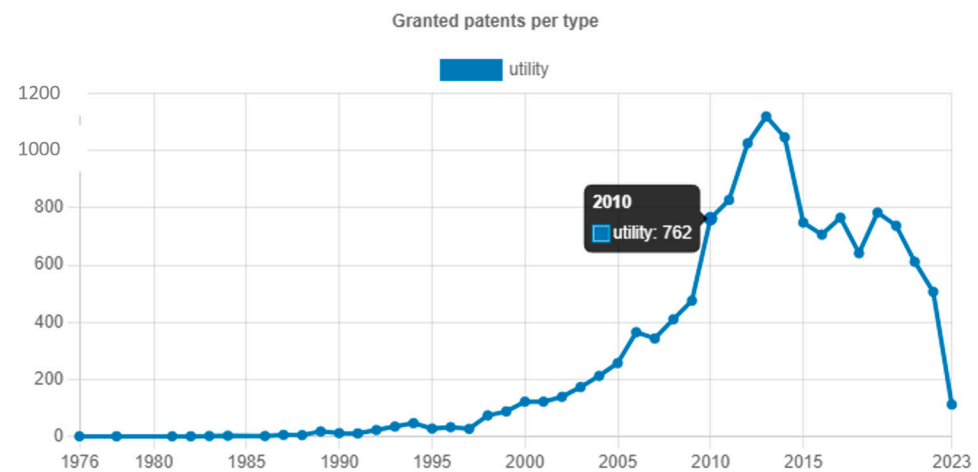


Figure 15. Granted patents per type for G06Q10/06.

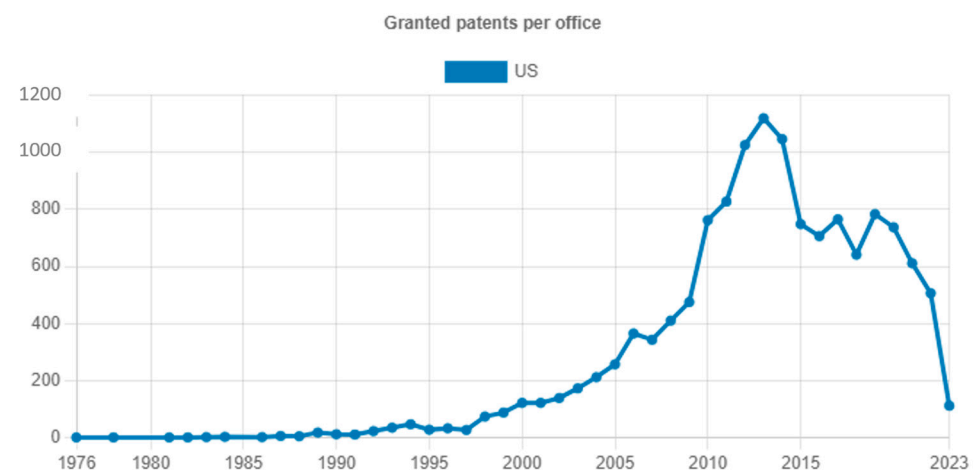


Figure 16. Granted patents per office for G06Q10/06.

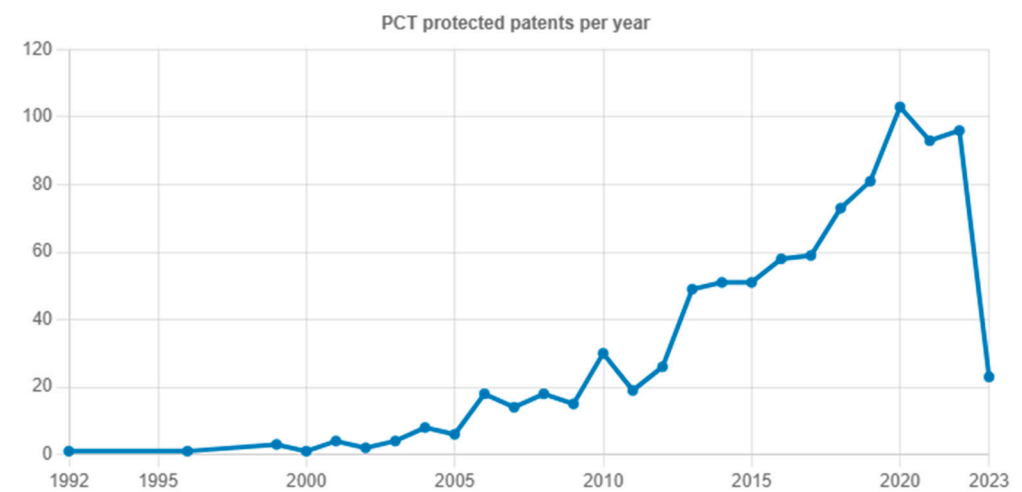


Figure 17. PCT granted patents for G06Q10/06.

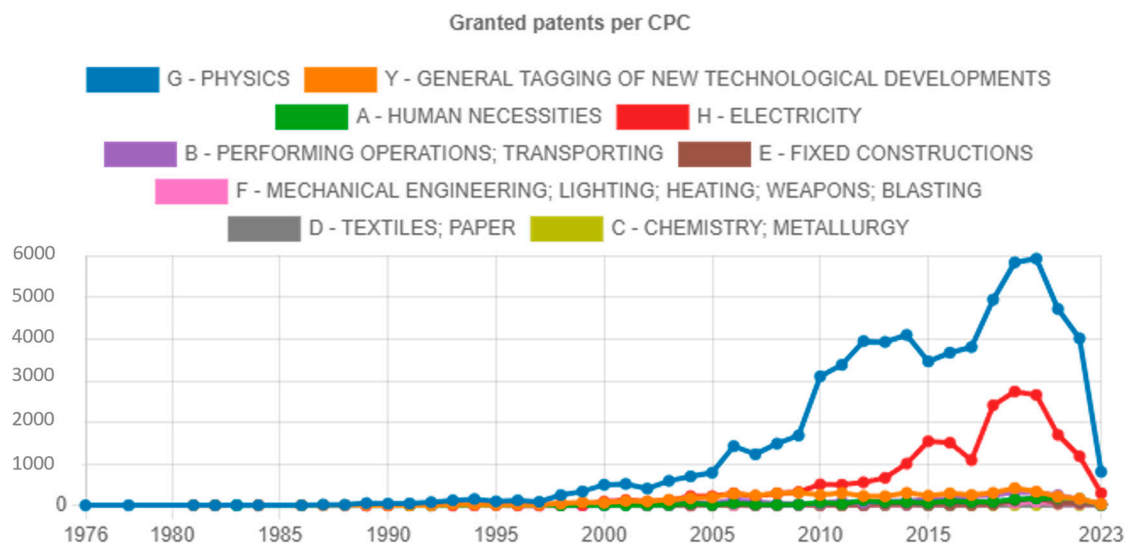


Figure 18. Granted patents per CPC section for G06Q10/06.

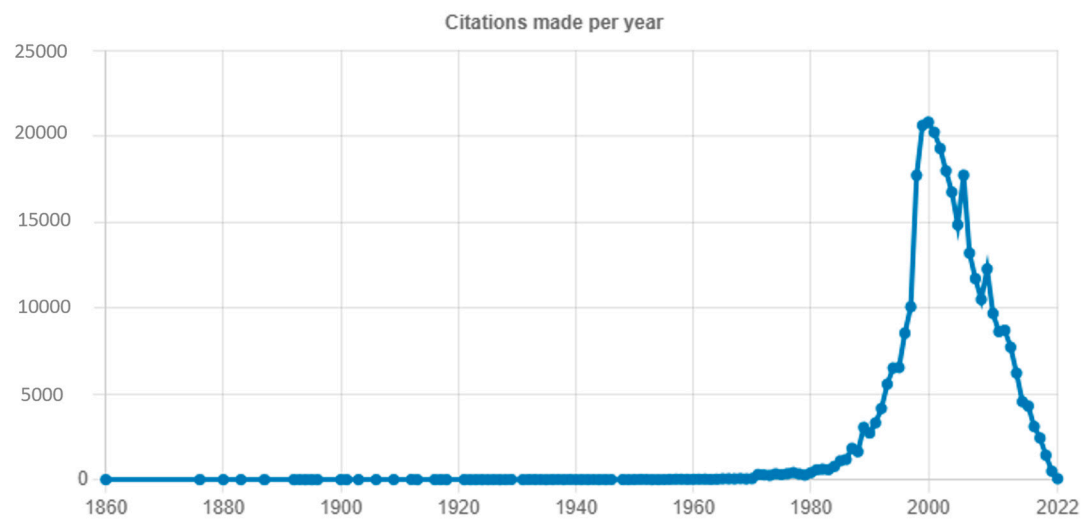


Figure 19. Citations made per year for G06Q10/06.

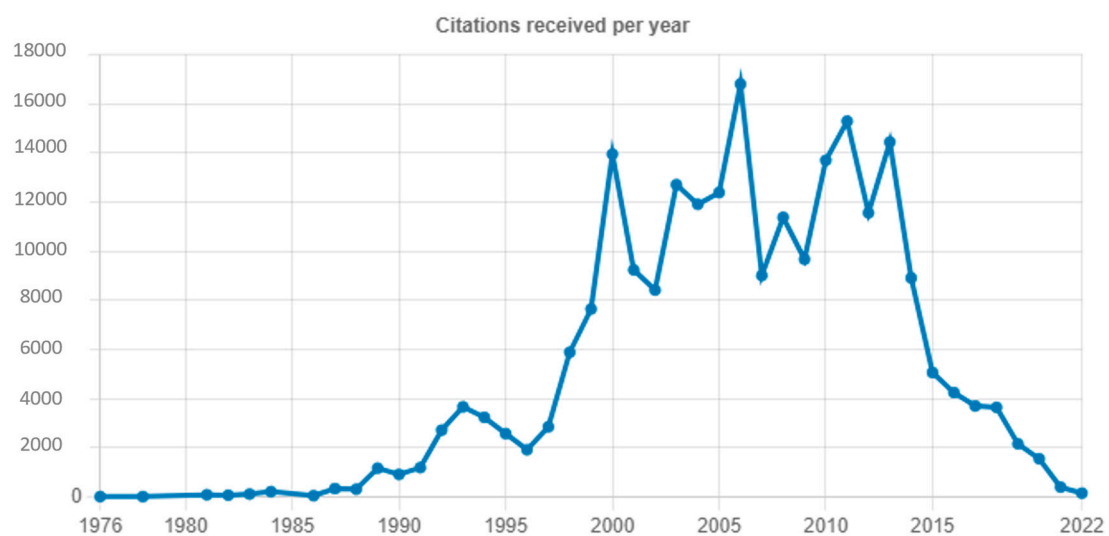
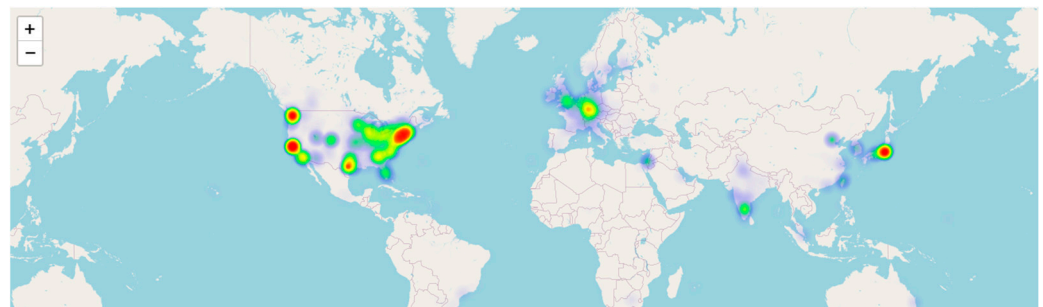


Figure 20. Citations received per year for G06Q10/06.

Table 8. Most productive inventors for G06Q10/06.

Inventor	Invention Count
Rick A. Hamilton, II	52
Curtis Chambers	46
Steven Nielsen	46
Jeffrey Farr	38
Kabir A. Barday	33
Clarence T. Tegreene	31
Robert W. Lord	29
Royce A. Levien	27
Gregory J. Boss	27
Edward B. Kalin	27

**Figure 21.** Inventor location distribution for G06Q10/06. Different colors represent the number of inventors in a region, with red indicating more inventors and green indicating less inventors.

The G06Q10/06 class is predominantly populated by Cooperation/Organization entities, making up 99.4% of the total. Among them, the leading assignees, presented in Table 9, tend to hold at least 124 patents each in the G06Q10/06 category. In total, the top 10 assignees collectively possess 3363 patents, accounting for approximately 25% of all patents within the G06Q10/06 category. The locations of the assignees, shown in Figure 22, closely mirror those of the inventors, but with a greater concentration in the prominent tech hubs across the United States, which is once again a result of USPTO being the only source of data. Among them, several reputable companies and organizations are visible, which employ a large pool of personnel and manage projects and resources on a complex scale, as well as companies that are involved in the Software and Informatics sectors such as IBM (USA), Microsoft (USA), Oracle (USA) and Accenture (IRL).

Table 9. Most prominent assignees for G06Q10/06.

Assignee	Patent Count
International Business Machines Corporation (US)	1508
Microsoft Technology Licensing, LLC (US)	418
SAP AG (DE)	333
Oracle International Corporation (US)	197
Hitachi, Ltd. (JP)	182
THE BOEING COMPANY (US)	160
Damian Wasserbauer General Electric Company (US)	156
Hewlett-Packard Development Company, L.P. (US)	151
Accenture Global Services Limited (IRL)	134
SAP SE (DE)	124

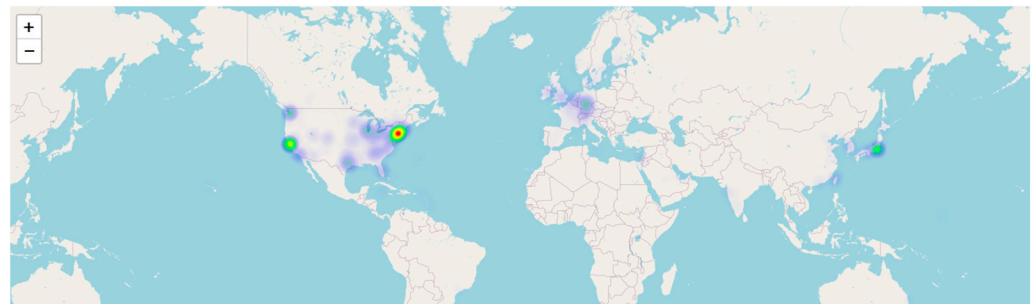


Figure 22. Assignee location distribution for G06Q10/06. Different colors represent the number of assignees in a region, with red indicating more assignees and green indicating less assignees.

PatentInspector was also employed to perform the LDA method for topic modeling, deleting the 20 most frequently appearing words in the document, using the default parameter of 10 topics. The execution of LDA yielded a coherence score of 0.523, meaning that the resulting topics were well-rounded and broad [25]. Subsequently, PatentInspector categorized these topics according to their patent share and the CAGR of their patent share during the period between 30 March 2015 and 28 March 2020, which covered 5 years before the last grant date to the last grant date itself, using the default settings of PatentInspector. Of course, the user has the capability of modifying any of these parameters according to his/her preference. It should be noted that this is an indicative, demonstrative execution of LDA, for the purposes of the case study, and not necessarily the optimal model. However, based on the outputs of the tool, we can provide some interpretations for the HRM field.

In Table 10, the extracted topics are presented, along with the top words of each topic and their patent share and CAGR classification. A title has also been assigned to each topic based on the words that characterize it and by inspecting the most representative patents that belong to it. In Table 11, the most representative patents for each topic are also included, as extracted by PatentInspector, aiding the user to validate the produced topics and assign titles to each topic, in combination with the most probable words.

Table 10. The 10 topics and their classification for G06Q10/06.

Title ¹	Patent Count	Words		Share	CAGR	Class
Resource Allocation and Supply Chain Analysis	838	Word	Weight	0.06	0.04	Emerging
		power	0.036			
		control	0.026			
		energy	0.026			
		resource	0.025			
		supply	0.018			
		load	0.014			
		plant	0.012			
		allocation	0.012			
		consumption	0.012			
Business Intelligence Software and Logic Diagrams	2450	usage	0.012	0.18	−0.01	Saturated
		Word	Weight			
		business	0.031			
		application	0.023			
		project	0.02			
		object	0.018			
		component	0.016			
		software	0.014			
		model	0.014			
		enterprise	0.011			
		configuration	0.01			
		tool	0.009			

Table 10. Cont.

Title ¹	Patent Count	Words		Share	CAGR	Class
Interface Interaction and Electronic Records	954	Word	Weight	0.07	0.02	Emerging
		display	0.029			
		interface	0.019			
		image	0.014			
		medical	0.013			
		graphical	0.012			
		patient	0.012			
		document	0.011			
		element	0.009			
		displayed	0.009			
		record	0.008			
Logic Programming and Event Handling	1318	Word	Weight	0.1	0.02	Emerging
		task	0.05			
		workflow	0.041			
		activity	0.02			
		event	0.019			
		node	0.019			
		rule	0.014			
		state	0.014			
		action	0.013			
		document	0.012			
		set	0.001			
Computing and IT Support	1185	Word	Weight	0.09	0.04	Emerging
		server	0.05			
		network	0.025			
		communication	0.019			
		client	0.009			
		configure	0.009			
		processor	0.008			
		second	0.008			
		computing	0.008			
		tag	0.008			
		identifier	0.007			
Job Scheduling and Product Ordering	1318	Word	Weight	0.1	−0.05	Declining
		product	0.035			
		production	0.027			
		order	0.024			
		manufacturing	0.021			
		job	0.019			
		schedule	0.017			
		supply	0.014			
		plan	0.014			
		planning	0.014			
		demand	0.013			
Autonomous Vehicles, Logistics and Tracking Operations	991	Word	Weight	0.07	0.03	Emerging
		vehicle	0.039			
		location	0.032			
		asset	0.03			
		mobile	0.011			
		repair	0.009			
		inspection	0.009			
		tracking	0.007			
		operation	0.007			
		autonomous	0.007			
		driver	0.006			

Table 10. Cont.

Title ¹	Patent Count	Words		Share	CAGR	Class
Hardware Maintenance and Processing	896	Word	Weight	0.07	−0.01	Declining
		unit	0.043			
		item	0.036			
		processing	0.025			
		work	0.023			
		apparatus	0.021			
		storage	0.015			
		equipment	0.013			
		maintenance	0.011			
		terminal	0.011			
		sensor	0.01			
Risk Assessment and Performance Evaluation metrics	1873	Word	Weight	0.14	−0.01	Saturated
		value	0.023			
		model	0.015			
		performance	0.013			
		set	0.012			
		analysis	0.011			
		determining	0.008			
		parameter	0.007			
		problem	0.007			
		risk	0.007			
		score	0.006			
Networking and Client Communication	1601	Word	Weight	0.12	0.01	Dominant
		message	0.015			
		request	0.014			
		network	0.013			
		content	0.01			
		customer	0.01			
		agent	0.01			
		access	0.009			
		event	0.009			
		notification	0.009			
		transaction	0.008			

¹ Titles are not generated by PatentInspector and are assigned based on user judgement.

Table 11. Most representative patents for each topic.

Topic 1	Topic 2
US10068020—Consumable data management	US10042904—System of centrally managing core reference data associated with an enterprise
US10360546—Method for supplying electrical power and billing for electrical power supplied using frequency regulation credits	US7310646—Data management system providing a data thesaurus for mapping between multiple data schemas or between multiple domains within a data schema
US6907381—System for aiding the preparation of operation and maintenance plans for a power-generation installation	US7885793—Method and system for developing a conceptual model to facilitate generating a business-aligned information technology solution
US8249756—Method, device and system for responsive load management using frequency regulation credits	US8204922—Master data management system for centrally managing core reference data associated with an enterprise
US9418046—Price-and-branch algorithm for mixed integer linear programming	US9021420—Deployment of business processes in service-oriented architecture environments

Table 11. Cont.

Topic 3	Topic 4
US10042524—Overview user interface of emergency call data of a law enforcement agency	US10032124—Hierarchical permissions model for case management
US10346010—Process data presentation based on process regions	US10296385—Dynamically modifying program execution capacity
US10387806—Digitizing venue maps	US10387153—Synchronizing a set of code branches
US10418131—System for providing identification and information, and for scheduling alerts	US10430253—Updating workflow nodes in a workflow
US10877638—Overview user interface of emergency call data of a law enforcement agency	US8381181—Updating a workflow when a user reaches an impasse in the workflow
Topic 5	Topic 6
US10785549—Technologies for switching network traffic in a data center	US5440480—Method for determining flexible demand in a manufacturing process
US7195149—Method of attaching an RF ID tag to a hose and tracking system	US6393332—Method and system for providing sufficient availability of manufacturing resources to meet unanticipated demand
US7436303—Rack sensor controller for asset tracking	US7587327—Order scheduling system and method for scheduling appointments over multiple days
US9112868—Client device, information processing system and associated methodology of accessing networked services	US9377476—Consumable data management
US9806891—System and method for an extended web of trust	US8204922—Master data management system for centrally managing core reference data associated with an enterprise
Topic 7	Topic 8
US10065653—Method and system for automatically identifying a driver by creating a unique driver profile for a vehicle from driving habits	US11288099—Electronic apparatus, storage medium storing computer program, and method of performing settings of electronic apparatus
US7555378—Driver activity and vehicle operation logging and reporting	US6965833—System and method for providing environmental impact information, recording medium recording the information, and computer data signal
US9104990—Article vending machine and method for exchanging an inoperable article for an operable article	US7536239—Chemical substance total management system, storage medium storing chemical substance management program and chemical substance total management method
US10685401—Communication of insurance claim data	US8200523—Procedure generation apparatus and method
US7327286—Marine vessel monitoring system	US7831659—Data providing system, server and program
Topic 9	Topic 10
US11538237—Utilizing artificial intelligence to generate and update a root cause analysis classification model	US10277556—Domain name hi-jack prevention
US5153366—Method for allocating and assigning defensive weapons against attacking weapons	US7796023—Systems and methods for the automatic registration of devices
US7643972—Computer-implemented systems and methods for determining steady-state confidence intervals	US7836482—Information management system
US7809781—Determining a time point corresponding to change in data values based on fitting with respect to plural aggregate value sets	US9159099—Exception notification system and method
US8001166—Methods and apparatus for optimizing keyword data analysis	US9191277—Method of registering a device at a remote site featuring a client application capable of detecting the device and transmitting registration messages between the device and the remote site

Among the detected topics, Resource Allocation and Supply Chain Analysis (Topic 1) is observed, as well as Risk Assessment & Performance Evaluation metrics (Topic 9), Job Scheduling (Topic 6) and the Analysis of Risk in Decisions (Topic 4). Moreover, some niche topics are also present, such as Computing & IT Support (Topic 5), Hardware Maintenance (Topic 8) and Network-Client Communication (Topic 10). Topics related to software, such as Business Intelligence software (Topic 2), which accelerates and simplifies HRM tasks, as well as Interface Interaction & Electronic Records (Topic 3) are also present. Finally, dealing with ways to automate logistics using autonomous vehicle technologies and tracking methods is also observed (Topic 7).

Based on the interpretation of the CAGR and patent share, extracted by PatentInspector, several trends have emerged in the field of HRM. Notably, areas such as logistics, operational tracking, supply chain analysis, logic programming, interface interaction and IT support are witnessing growth. Conversely, domains like networking and client communication remain predominant. In contrast, job scheduling and hardware maintenance show signs of decline. Additionally, business intelligence as well as risk assessment and evaluation are currently exhibiting a state of saturation.

In the G06Q10/06 local citation network consisting of 27,702 citations (Table 12), the majority of highly cited patents have garnered no less than 142 citations. These patents primarily focus on decision support, resource management systems and associated methodologies, while job scheduling, optimization and sales automation also make an appearance, linking the cited patents with the extracted topics from the topic analysis component. Meanwhile, within the G06Q10/06 global citation network, which measures the citations from the entire USPTO database, the most cited patents have a minimum of 780 citations, as highlighted in Table 13, with a predominant emphasis on networking, client communications, collaboration and resource management. We can see that while local patents are more specific and targeted in their objectives, the global patents are more abstract in their purposes, which is expected when considering that patents from this class may be used as citations from other patents of different domains and may hence concern other concepts.

Table 12. Most cited patents on the local network for G06Q10/06.

Patent	Incoming Citations
US6151582—Decision support system for the management of an agile supply chain	142
US5953707—Decision support system for the management of an agile supply chain	138
US5630070—Optimization of manufacturing resource planning	125
US4937743—Method and system for scheduling, monitoring and dynamically managing resources	117
US6578005—Method and apparatus for resource allocation when schedule changes are incorporated in real time	110
US5369570—Method and system for continuous integrated resource management	109
US5826239—Distributed workflow resource management system and method	108
US5189606—Totally integrated construction cost estimating, analysis, and reporting system	107
US5111391—System and method for making staff schedules as a function of available resources as well as employee skill level, availability and priority	92
US5216612—Intelligent computer integrated maintenance system and method	90

The comparison of the constructed case study of PatentInspector with a replication in Lens (which can be found in the Supplementary Materials—Case Study #1) indicates that the extracted inventors, assignees and globally cited patents along with the timeline of granted patents correspond with the insights from PatentInspector. Hence, the alignment of the constructed tool with an established source of patent data and descriptive PA is an encouraging indicator of the validity of PatentInspector and its potential for scientific PA.

Table 13. Most cited patents on the global network for G06Q10/06.

Patent	Incoming Citations
US US6850895—Assignment manager	2091
US6665648—State models for monitoring process	1932
US8082301—System for supporting collaborative activity	1470
US7356482—Integrated change management unit	1263
US8484111—Integrated change management unit	1238
US6835173—Robotic endoscope	1144
US6770027—Robotic endoscope with wireless interface	1093
US7124101—Asset tracking in a network-based supply chain environment	905
US6671818—Problem isolation through translating and filtering events into a standard object format in a network based supply chain	783
US8468244—Digital information infrastructure and method for security designated data and with granular data stores	780

6. Discussion and Implications

Based on the insights derived from the investigated case study of Section 5, we can deduce that the use of PatentInspector facilitated the interpretation of HRM patents and profiles HRM as an active field, with an abundance of patent applications and grants, particularly in the last 10 years. Based on the topics extracted by the Topic Analysis Tab and the inspection of the most representative documents provided by the tool, it appears that the emergence of software solutions and the constant provision of data have certainly influenced the topics, objectives and purposes of patents in this field, and many reputable organizations have been granted patents related to HRM.

The use of PatentInspector showcases that HRM patents have a mean granting time of 3.6 years, among other useful statistics produced from the Descriptive Analysis Tab. The profiling of active inventors and assignees indicated that companies such as IBM, Microsoft and Amazon are interested in HRM patents, while PatentInspector provides an overview of their locations and the evolution of several variables over time (e.g., number of citations).

The analysis conducted by the Topic Analysis Tab profiled the primary objectives of HRM patents, presenting the status of the topics and the most representative patents. An observation of the extracted topics by the LDA methodology presents autonomous vehicles and logistics as emerging topics, along with hardware maintenance. In general, the Topic Analysis Tab provided an easy means of assessing the primary trends in HRM patents and whether these trends dominate or have saturated the market, using the implemented CAGR metric. Overall, the executed LDA model is well rounded, with a good coherence score, indicating that the extracted topics capture the semantics and objectives of HRM in a concise manner.

Moreover, the Network Analysis Tab portrays the most cited (locally or globally) patent entries, allowing users to view which patents are more influential among the retrieved documents and examine which technologies or patent objectives may shape or influence subsequent patent applications.

Based on the results of the case study, we can deduce that PatentInspector is an easy-to-use and practical tool for PA, with the core insights produced by the tool providing the potential to assess the developments in a patent domain, with an emphasis on the US. The tool fully portrays the most prolific organizations, inventors and locations, while also being able to showcase the primary topics of patent objectives and their growth in a given time period. Moreover, the citation networks allow users to examine which patents are more popular among patent applications and are consistently used as reference points. This process is achieved via the use of streamlined visualizations that facilitate user understanding and require little or no scientific and coding background to be interpreted.

Evidently, PatentInspector serves as an easy-to-learn, public resource that, while not being able to replace more complex PA methodologies, can certainly facilitate the carrying out of basic PA tasks, while also offering opportunities for some higher-level analysis, such

as topic modeling with two established algorithms. The simplicity of the tool encourages users of different backgrounds, ranging from PA enthusiasts to seasoned researchers, to leverage its capabilities and perform a baseline analysis for a patent domain of their choice. Moreover, the tool is not proprietary and is already deployed and ready-to-use, while the codebase is fully open-source and extensible.

7. Threats to Validity

In this section, we present some threats to the validity of the proposed PA software (v.1), making the distinction between internal validity, i.e., limitations in the methodological design of PatentInspector, and external validity, i.e., factors that limit the generalization and applicability of PatentInspector to other domains or patent offices.

Regarding internal validity, one primary limitation is that PatentInspector only retrieves data from the USPTO and not from other major patent offices, such as the EPO or CNIPA. This automatically skews the results, as the illustrated plots, topics and citation networks will inevitably present a partial view of patent grants, with a focus on the US region. However, this threat is mitigated by the fact that the USPTO has been indicated in the literature to be a viable source that effectively captures global patent trends [111,117]. It should also be emphasized that the choice of the USPTO was based on the fact that it was the only patent office to include a “bulk data” endpoint that could allow the storage of the entire patent office in the database of PatentInspector, given the resources and limitations to access and real-time data retrieval when developing the tool. We recognize that this is a threat and plan to expand the tool to include more data sources in subsequent versions.

In addition, the developed tool leverages data from patents that have already received a grant and does not consider patents that have been applied for and are pending evaluation. While this may lead to data omission, it is a reasonable practice, as applied patents may be rejected, in contrast to granted patents that have been carefully examined. In addition, a minor threat to the developed tool is that we do not introduce new methodologies for PA or leverage advanced methodologies for strategic analysis, technology convergence or business scoping. However, as our primary goal was to introduce a PA tool accessible to multiple parties with various backgrounds, the features that were chosen and incorporated focused on scientific concepts that can be easily understood and interpreted by individuals of different levels.

In terms of our reliance on user judgement and expertise, this only has relevance when experimenting with the implemented topic analysis algorithms. An experienced user can alter the values of topic modeling (number of models, keywords to be removed, time range) and experiment with different setups. However, we consider this a minor threat, as the topic modeling aspect of PatentInspector is fully supported to run on the default parameters and produce reliable results. Finally, as far as the waiting times for the analysis to be conducted are concerned, this is indeed a limitation in the functionalities of PatentInspector, as the deployed server cannot support a large number of simultaneous users. To mitigate this, we have implemented an alert function that allows users to exit the tool, run the analysis in the background and receive an email once the report is generated.

Regarding the external validity of PatentInspector, the applicability of our data retrieval, preprocessing, storage and analysis could potentially extend to other patent offices but it may be hindered both by closed and proprietary APIs, as well as possible different patenting procedures, which may lead to different data being stored. Thus, any application of our tool to another office, such as the EPO, should be carefully structured, with proper adjustments to the database in order to accommodate potentially varied patent data. In addition, as PatentInspector relies on predownloaded data from USPTO that are periodically updated, it cannot be used as a patent database that retrieves data “on the fly”, but rather as a tool for analysis that utilizes the most recent snapshot of patent data from the USPTO. Finally, PatentInspector focuses only on the scientific aspect of PA, demonstrating useful statistics, topics and citations, and does not delve into the legal procedures of patenting, such as litigations, or refined economic indicators. However, this is not a major threat as

our primary goal was to offer an open-source scientific software application that mainly targets PA researchers and scientists, but which could also be used by industrial actors as a complementary tool for the analysis of trends, in conjunction with appropriate business intelligence suites.

8. Contributions and Conclusions

8.1. Contributions

In this study, our vision was to offer an application of this scope, creating a flexible tool for PA that is open-source, free to use and provides interpretable insights for multiple interested parties.

The developed tool can indeed be used to extract descriptive statistics, thematic axes and citation analysis, focusing on the USPTO and being capable of analyzing thousands of patent records. Its usability was demonstrated in a case study of HRM patents, where the extracted visualizations captured the landscape of the domain and allowed the rapid detection of active inventors, prolific organizations and emerging or dominant thematic axes.

Overall, the tool that we offer contributes to the current landscape of PA tools by (i) offering a publicly deployed, direct and easy-to-use solution for PA that can be used by users without coding or advanced PA knowledge; (ii) providing a Topic Modeling panel that can be used by researchers to extract thematic axes on patent data while also evaluating the growth or decline of each topic; (iii) producing flexible visualizations that can be easily interpreted by all users, without requiring advanced background knowledge of PA; and (iv) having the source code of the tool publicly available and open-source, to be modified or improved by any researcher that wishes to expand the tool's functionalities.

We believe that PatentInspector is a valuable resource for any individual that wishes to conduct a baseline PA study, without being limited by pricing or knowledge gaps.

8.2. Conclusions and Future Work

The field of PA is evolving rapidly, being applied to a plethora of domains for various different objectives. The abundance of patent data and the constant need for analysis has led to a range of tools and software that facilitate this purpose. Especially due to the rise of open science and scientific software development, applications and tools that encourage scientists to openly engage with software and advance their research are more than necessary. We believe that bridging the field of PA with the open-source community can yield multiple benefits for all interested parties, advancing research and scientific maturity and promoting easy access to knowledge and learning.

Some future work directions of this study include expanding our database to include patents from other offices, with the EPO being an important source, as well as configuring our patent database to periodically be updated and also include patent families (single or extended), using the latest data from the USPTO. In addition, we plan to enhance the capabilities of PatentInspector by adding more advanced methodologies for topic modeling, along with a grid search function to find the optimal model, technological convergence (e.g., convergence networks) and co-word analysis [118]. Finally, linking PatentInspector with existing patent databases that could enable the faster retrieval of data would greatly accelerate the storage process and would elevate the user experience.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/app132413147/s1>. The source code of PatentInspector can be downloaded from <https://github.com/KonstantinosPetrakis/PatentInspector/tree/mdpi> (accessed on 8 December 2023), while the tool is publicly available for use at <https://patentinspector.csd.auth.gr/> (accessed on 8 December 2023).

Author Contributions: Conceptualization, K.P., K.G., N.M. and L.A.; Data curation, K.P.; Formal analysis, K.P. and K.G.; Investigation, K.P. and K.G.; Methodology, K.P., K.G., N.M. and L.A.; Project administration, L.A.; Software, K.P. and K.G.; Supervision, N.M. and L.A.; Validation, K.P.; Visualiza-

tion, K.P. and K.G.; Writing—original draft, K.P.; Writing—review and editing, K.P., K.G., N.M. and L.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The source code and data of the presented framework are open-source and available at <https://github.com/KonstantinosPetrakis/PatentInspector/tree/mdpi> (accessed on 8 December 2023).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. 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).
2. Lemley, M. Understanding the realities of modern patent litigation [Preprint]. *Tex. Law Rev.* **2016**, *92*, 1769. [CrossRef]
3. Abraham, B.P.; Moitra, S.D. Innovation assessment through patent analysis. *Technovation* **2001**, *21*, 245–252. [CrossRef]
4. 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. [CrossRef]
5. Tseng, Y.-H.; Lin, C.-J.; Lin, Y.-I. Text mining techniques for patent analysis. *Inf. Process. Manag.* **2007**, *43*, 1216–1247. [CrossRef]
6. Breitzman, A.F.; Mogee, M.E. The many applications of patent analysis. *J. Inf. Sci.* **2002**, *28*, 187–205. [CrossRef]
7. Foster, E.D.; Deardorff, A. Open science framework (OSF). *J. Med. Libr. Assoc. JMLA* **2017**, *105*, 203. [CrossRef]
8. Vicente-Saez, R.; Martinez-Fuentes, C. Open Science now: A systematic literature review for an integrated definition. *J. Bus. Res.* **2018**, *88*, 428–436. [CrossRef]
9. Segal, J.; Morris, C. Developing Scientific Software. *IEEE Softw.* **2008**, *25*, 18–20. [CrossRef]
10. Nguyen-Hoan, L.; Flint, S.; Sankaranarayanan, 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.
11. 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. [CrossRef]
12. Available online: <https://github.com/search?q=patent+analysis&type=repositories> (accessed on 29 October 2023).
13. 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. [CrossRef]
14. Fujii, H.; Managi, S. Trends and priority shifts in artificial intelligence technology invention: A global patent analysis. *Econ. Anal. Policy* **2018**, *58*, 60–69. [CrossRef]
15. Tseng, C.-Y.; Ting, P.-H. Patent analysis for technology development of artificial intelligence: A country-level comparative study. *Innovation* **2013**, *15*, 463–475. [CrossRef]
16. 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. [CrossRef]
17. 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. [CrossRef]
18. 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. [CrossRef]
19. 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. [CrossRef]
20. 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. [CrossRef]
21. 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. [CrossRef]
22. 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. [CrossRef]
23. Daim, T.U.; Rueda, G.; Martin, H.; Gerdri, P. Forecasting emerging technologies: Use of bibliometrics and patent analysis. *Technol. Forecast. Soc. Chang.* **2006**, *73*, 981–1012. [CrossRef]
24. 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. [CrossRef]
25. Blei, D.M.; Ng, A.Y.; Jordan, M.I. Latent dirichlet allocation. *J. Mach. Learn. Res.* **2003**, *3*, 993–1022.
26. Ghaffari, M.; Aliahmadi, A.; Khalkhali, A.; Zakery, A.; Daim, T.U.; Yalcin, H. Topic-based technology mapping using patent data analysis: A case study of vehicle tires. *Technol. Forecast. Soc. Chang.* **2023**, *193*, 122576. [CrossRef]

27. Zhang, W.; Cao, G.; Ji, Y.; Gu, L.; Wang, S. Analysis of electric vehicle technology development based on patent big data: A topic analysis of structured topic model (STM). In Proceedings of the 5th International Conference on Computer Information Science and Application Technology (CISAT 2022), Chongqing, China, 29–31 July 2022; Volume 12451, pp. 565–571.
28. Wang, H.; Zhang, D.; Zhai, C. Structural topic model for latent topical structure analysis. In Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, Portland, OR, USA, 19–24 June 2011; pp. 1526–1535.
29. Choi, H.; Woo, J. Investigating emerging hydrogen technology topics and comparing national level technological focus: Patent analysis using a structural topic model. *Appl. Energy* **2022**, *313*, 118898. [\[CrossRef\]](#)
30. Wang, J.; Hsu, C.-C. A topic-based patent analytics approach for exploring technological trends in smart manufacturing. *J. Manuf. Technol. Manag.* **2021**, *32*, 110–135. [\[CrossRef\]](#)
31. Kim, D.; Kim, S. Role and challenge of technology toward a smart sustainable city: Topic modeling, classification, and time series analysis using information and communication technology patent data. *Sustain. Cities Soc.* **2022**, *82*, 103888. [\[CrossRef\]](#)
32. Georgiou, K.; Mittas, N.; Ampatzoglou, A.; Chatzigeorgiou, A.; Angelis, L. Data-Oriented Software Development: The Industrial Landscape through Patent Analysis. *Information* **2022**, *14*, 4. [\[CrossRef\]](#)
33. Wang, B.; Liu, S.; Ding, K.; Liu, Z.; Xu, J. Identifying technological topics and institution-topic distribution probability for patent competitive intelligence analysis: A case study in LTE technology. *Scientometrics* **2014**, *101*, 685–704. [\[CrossRef\]](#)
34. Huang, L.; Hou, Z.; Fang, Y.; Liu, J.; Shi, T. Evolution of CCUS Technologies Using LDA Topic Model and Derwent Patent Data. *Energies* **2023**, *16*, 2556. [\[CrossRef\]](#)
35. Kim, G.J.; Park, S.S.; Jang, D.S. Technology forecasting using topic-based patent analysis. *J. Sci. Ind. Res.* **2015**, *74*, 265–270.
36. Kim, K.H.; Han, Y.J.; Lee, S.; Cho, S.W.; Lee, C. Text mining for patent analysis to forecast emerging technologies in wireless power transfer. *Sustainability* **2019**, *11*, 6240. [\[CrossRef\]](#)
37. Choi, D.; Song, B. Exploring technological trends in logistics: Topic modeling-based patent analysis. *Sustainability* **2018**, *10*, 2810. [\[CrossRef\]](#)
38. Kim, M.; Park, Y.; Yoon, J. Generating patent development maps for technology monitoring using semantic patent-topic analysis. *Comput. Ind. Eng.* **2016**, *98*, 289–299. [\[CrossRef\]](#)
39. Ma, T.; Zhou, X.; Liu, J.; Lou, Z.; Hua, Z.; Wang, R. Combining topic modeling and SAO semantic analysis to identify technological opportunities of emerging technologies. *Technol. Forecast. Soc. Chang.* **2021**, *173*, 121159. [\[CrossRef\]](#)
40. Zhang, H.; Daim, T.; Zhang, Y. Integrating patent analysis into technology roadmapping: A latent dirichlet allocation based technology assessment and roadmapping in the field of Blockchain. *Technol. Forecast. Soc. Chang.* **2021**, *167*, 120729. [\[CrossRef\]](#)
41. Chen, H.; Zhang, G.; Zhu, D.; Lu, J. Topic-based technological forecasting based on patent data: A case study of Australian patents from 2000 to 2014. *Technol. Forecast. Soc. Chang.* **2017**, *119*, 39–52. [\[CrossRef\]](#)
42. Kim, G.; Bae, J. A novel approach to forecast promising technology through patent analysis. *Technol. Forecast. Soc. Chang.* **2017**, *117*, 228–237. [\[CrossRef\]](#)
43. Suominen, A.; Toivanen, H.; Seppänen, M. Firms' knowledge profiles: Mapping patent data with unsupervised learning. *Technol. Forecast. Soc. Chang.* **2017**, *115*, 131–142. [\[CrossRef\]](#)
44. Momeni, A.; Rost, K. Identification and monitoring of possible disruptive technologies by patent-development paths and topic modeling. *Technol. Forecast. Soc. Chang.* **2016**, *104*, 16–29. [\[CrossRef\]](#)
45. Mariani, M.S.; Medo, M.; Lafond, F. Early identification of important patents: Design and validation of citation network metrics. *Technol. Forecast. Soc. Chang.* **2019**, *146*, 644–654. [\[CrossRef\]](#)
46. van Raan, A.F. Patent Citations Analysis and Its Value in Research Evaluation: A Review and a New Approach to Map Technology-relevant Research. *J. Data Inf. Sci.* **2017**, *2*, 13–50. [\[CrossRef\]](#)
47. Li, X.; Yuan, X. Tracing the technology transfer of battery electric vehicles in China: A patent citation organization network analysis. *Energy* **2022**, *239*, 122265. [\[CrossRef\]](#)
48. Lee, S.; Kim, W.; Lee, H.; Jeon, J. Identifying the structure of knowledge networks in the US mobile ecosystems: Patent citation analysis. *Technol. Anal. Strat. Manag.* **2016**, *28*, 411–434. [\[CrossRef\]](#)
49. Ferrari, V.E.; da Silveira, J.M.F.J.; Dal-Poz, M.E.S. Patent network analysis in agriculture: A case study of the development and protection of biotechnologies. *Econ. Innov. New Technol.* **2021**, *30*, 111–133. [\[CrossRef\]](#)
50. Mao, G.; Han, Y.; Liu, X.; Crittenden, J.; Huang, N.; Ahmad, U.M. Technology status and trends of industrial wastewater treatment: A patent analysis. *Chemosphere* **2022**, *288*, 132483. [\[CrossRef\]](#) [\[PubMed\]](#)
51. Ji, J.; Barnett, G.A.; Chu, J. Global networks of genetically modified crops technology: A patent citation network analysis. *Scientometrics* **2019**, *118*, 737–762. [\[CrossRef\]](#)
52. Choe, H.; Lee, D.H.; Kim, H.D.; Seo, I.W. Structural properties and inter-organizational knowledge flows of patent citation network: The case of organic solar cells. *Renew. Sustain. Energy Rev.* **2016**, *55*, 361–370. [\[CrossRef\]](#)
53. Choe, H.; Lee, D.H.; Seo, I.W.; Kim, H.D. Patent citation network analysis for the domain of organic photovoltaic cells: Country, institution, and technology field. *Renew. Sustain. Energy Rev.* **2013**, *26*, 492–505. [\[CrossRef\]](#)
54. You, H.; Li, M.; Hipel, K.W.; Jiang, J.; Ge, B.; Duan, H. Development trend forecasting for coherent light generator technology based on patent citation network analysis. *Scientometrics* **2017**, *111*, 297–315. [\[CrossRef\]](#)
55. Kim, E.; Cho, Y.; Kim, W. Dynamic patterns of technological convergence in printed electronics technologies: Patent citation network. *Scientometrics* **2014**, *98*, 975–998. [\[CrossRef\]](#)

56. Li, X.; Chen, H.; Huang, Z.; Roco, M.C. Patent citation network in nanotechnology (1976–2004). *J. Nanopart. Res.* **2007**, *9*, 337–352. [CrossRef]
57. Park, Y.-N.; Lee, Y.-S.; Kim, J.-J.; Lee, T.S. The structure and knowledge flow of building information modeling based on patent citation network analysis. *Autom. Constr.* **2018**, *87*, 215–224. [CrossRef]
58. Chang, S.B.; Lai, K.K.; Chang, S.M. Exploring technology diffusion and classification of business methods: Using the patent citation network. *Technol. Forecast. Soc. Chang.* **2009**, *76*, 107–117. [CrossRef]
59. Cho, T.-S.; Shih, H.-Y. Patent citation network analysis of core and emerging technologies in Taiwan: 1997–2008. *Scientometrics* **2011**, *89*, 795–811. [CrossRef]
60. Huang, Y.; Li, R.; Zou, F.; Jiang, L.; Porter, A.L.; Zhang, L. Technology life cycle analysis: From the dynamic perspective of patent citation networks. *Technol. Forecast. Soc. Chang.* **2022**, *181*, 121760. [CrossRef]
61. Érdi, P.; Makovi, K.; Somogyvári, Z.; Strandburg, K.; Tobochnik, J.; Volf, P.; Zálányi, L. Prediction of emerging technologies based on analysis of the US patent citation network. *Scientometrics* **2013**, *95*, 225–242. [CrossRef]
62. Ji, Y.; Yu, X.; Sun, M.; Zhang, B. Exploring the Evolution and Determinants of Open Innovation: A Perspective from Patent Citations. *Sustainability* **2022**, *14*, 1618. [CrossRef]
63. Fontana, R.; Nuvolari, A.; Verspagen, B. Mapping technological trajectories as patent citation networks. An application to data communication standards. *Econ. Innov. New Technol.* **2009**, *18*, 311–336. [CrossRef]
64. Kumar, V.; Lai, K.-K.; Chang, Y.-H.; Lin, C.-Y. Mapping technological trajectories for energy storage device through patent citation network. In Proceedings of the 2018 9th International Conference on Awareness Science and Technology (iCAST), Fukuoka, Japan, 19–21 September 2018; pp. 56–61.
65. Verspagen, B. Mapping technological trajectories as patent citation networks: A study on the history of fuel cell research. Advances in complex systems. *Adv. Complex Syst.* **2007**, *10*, 93–115. [CrossRef]
66. Yu, D.; Pan, T. Identifying technological development trajectories in blockchain domain: A patent citation network analysis. *Technol. Anal. Strat. Manag.* **2021**, *33*, 1484–1497. [CrossRef]
67. von Wartburg, I.; Teichert, T.; Rost, K. Inventive progress measured by multi-stage patent citation analysis. *Res. Policy* **2005**, *34*, 1591–1607. [CrossRef]
68. Wang, X.; Daim, T.; Huang, L.; Li, Z.; Shaikh, R.; Kassi, D.F. Monitoring the development trend and competition status of high technologies using patent analysis and bibliographic coupling: The case of electronic design automation technology. *Technol. Soc.* **2022**, *71*, 102076. [CrossRef]
69. Hu, X.; Rousseau, R.; Chen, J. A new approach for measuring the value of patents based on structural indicators for ego patent citation networks. *J. Am. Soc. Inf. Sci. Technol.* **2012**, *63*, 1834–1842. [CrossRef]
70. Yang, G.-C.; Li, G.; Li, C.-Y.; Zhao, Y.-H.; Zhang, J.; Liu, T.; Chen, D.-Z.; Huang, M.-H. Using the comprehensive patent citation network (CPC) to evaluate patent value. *Scientometrics* **2015**, *105*, 1319–1346. [CrossRef]
71. Chakraborty, M.; Byshkin, M.; Crestani, F. Patent citation network analysis: A perspective from descriptive statistics and ERGMs. *PLoS ONE* **2020**, *15*, e0241797. [CrossRef]
72. Gould, R.V.; Fernandez, R.M. Structures of mediation: A formal approach to brokerage in transaction networks. *Sociol. Methodol.* **1989**, *19*, 89. [CrossRef]
73. Suh, Y.; Jeon, J. Monitoring patterns of open innovation using the patent-based brokerage analysis. *Technol. Forecast. Soc. Chang.* **2019**, *146*, 595–605. [CrossRef]
74. Reymond, D.; Quoniam, L. A new patent processing suite for academic and research purposes. *World Pat. Inf.* **2016**, *47*, 40–50. [CrossRef]
75. A Patent Collector and Analyser to Expand Your Horizon with Various Data Processing Tools for Education and Scientific Purposes. Available online: <http://patent2netv2.vlab4u.info/> (accessed on 29 October 2023).
76. Tang, J.; Wang, B.; Yang, Y.; Hu, P.; Zhao, Y.; Yan, X.; Gao, B.; Huang, M.; Xu, P.; Li, W.; et al. Patentminer: Topic-driven patent analysis and mining. In Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Beijing, China, 12–16 August 2012; pp. 1366–1374.
77. Wu, Y.; Ji, Y.; Gu, F.; Guo, J. A collaborative evaluation method of the quality of patent scientific and technological resources. *World Pat. Inf.* **2021**, *67*, 102074. [CrossRef]
78. Perez-Molina, E.; Loizides, F. Novel data structure and visualization tool for studying technology evolution based on patent information: The DTFootprint and the TechSpectrogram. *World Pat. Inf.* **2021**, *64*, 102009. [CrossRef]
79. Patent Search and Analysis Database. PatSeer. 2023. Available online: <https://patseer.com/> (accessed on 29 October 2023).
80. Derwent Innovations Index on Web of Science. Clarivate. 2023. Available online: <https://clarivate.com/products/scientific-and-academic-research/research-discovery-and-workflow-solutions/webofscience-platform/derwent-innovations-index-on-web-of-science/> (accessed on 29 October 2023).
81. Orbit Intelligence—Patent Search & Analytics Software (No Date) Questel. Available online: <https://www.questel.com/patent/ip-intelligence-software/orbit-intelligence/> (accessed on 29 October 2023).
82. Patent Platform. Available online: <https://iamip.com/> (accessed on 29 October 2023).
83. AI Patent Search & Classification. Available online: <https://www.iprally.com/> (accessed on 29 October 2023).
84. PatBase—Leading Online Patent Search Database. Available online: <https://minesoft.com/patbase/> (accessed on 29 October 2023).

85. Deter Abuse. Available online: <https://www.unifiedpatents.com/> (accessed on 29 October 2023).
86. Art, S.P. Patent search, technology landscaping, ML Tools & Building Big Data Engineering. Available online: <https://www.patent-art.com/> (accessed on 29 October 2023).
87. Commercialize More IP. Available online: <https://www.tradespace.io/> (accessed on 29 October 2023).
88. Patent Analytics & Search Software. Available online: <https://www.acclaimip.com/> (accessed on 29 October 2023).
89. Innography IP Intelligence Software. Available online: <https://clarivate.com/products/ip-intelligence/patent-intelligence-software/innography/> (accessed on 29 October 2023).
90. IPlytics. Available online: <https://www.lexisnexisip.com/solutions/ip-analytics-and-intelligence/iplytics/> (accessed on 29 October 2023).
91. Minesoft Origin: Advanced AI Patent Search. Available online: <https://minesoft.com/minesoft-origin/> (accessed on 29 October 2023).
92. Octimine. Available online: <https://www.dennemeyer.com/octimine/> (accessed on 29 October 2023).
93. Patent Inspiration. Available online: <https://www.patentinspiration.com/> (accessed on 29 October 2023).
94. PatentSight. Available online: <https://www.lexisnexisip.com/solutions/ip-analytics-and-intelligence/patentsight/> (accessed on 29 October 2023).
95. Patsnap. Available online: <https://www.patsnap.com/> (accessed on 29 October 2023).
96. Patent Analysis Software—Patent Insight Pro. Available online: <https://www.patentinsightpro.com/> (accessed on 29 October 2023).
97. Patent Quality Artificial Intelligence. Available online: <https://projectpq.ai/> (accessed on 29 October 2023).
98. Ip-Tools IP-Tools/Patzilla: PatZilla Is a Modular Patent Information Research Platform and Data Integration Toolkit with a Modern User Interface and Access to Multiple Data Sources. Available online: <https://github.com/ip-tools/patzilla> (accessed on 29 October 2023).
99. Google Patents. Available online: <https://patents.google.com/> (accessed on 29 October 2023).
100. Available online: <https://www.freepatentonline.com/login.html> (accessed on 29 October 2023).
101. Where AI Can Make a Difference to the Innovation Ecosystem. Available online: <https://relecura.com/> (accessed on 29 October 2023).
102. The Lens—Free & Open Patent and Scholarly Search. Available online: <https://www.lens.org/> (accessed on 29 October 2023).
103. Kamilien1/Patentr: Patent Analysis Tool in R. Available online: <https://github.com/kamilien1/patentr> (accessed on 29 October 2023).
104. SumoBrain—Big, Powerful, Smart Searching. Available online: <https://www.sumobrain.com/> (accessed on 29 October 2023).
105. Patent Analyzer (No Date) PatentPC. Available online: <https://patentpc.com/patent-analyzer/> (accessed on 29 October 2023).
106. Data Driven Patent Analytics Tool: Patexia Patent Analyzer (No Date) Data Driven Patent Analytics Tool | Patexia Patent Analyzer. Available online: <https://www.patexia.com/insight/prosecution-analyzer> (accessed on 29 October 2023).
107. Django. Available online: <https://www.djangoproject.com/> (accessed on 15 March 2015).
108. The Progressive Javascript Framework. Available online: <https://vuejs.org/> (accessed on 29 October 2023).
109. Oliphant, T.E. Python for scientific computing. *Comput. Sci. Eng.* **2007**, *9*, 10–20. [CrossRef]
110. Data Download Tables. Available online: <https://patentsview.org/download/data-download-tables> (accessed on 29 October 2023).
111. Kim, J.; Lee, S. Patent databases for innovation studies: A comparative analysis of USPTO, EPO, JPO and KIPO. *Technol. Forecast. Soc. Chang.* **2015**, *92*, 332–345. [CrossRef]
112. Lee, D.; Seung, H.S. Algorithms for non-negative matrix factorization. In Proceedings of the 13th International Conference on Neural Information Processing Systems, Denver, CO, USA, 1 January 2000; MIT Press: Cambridge, MA, USA, 2000; pp. 535–541.
113. Tay, A. 6 Reasons Why you Should Try lens.org, Medium. 2019. Available online: <https://aarontay.medium.com/6-reasons-why-you-should-try-lens-org-c40abb09ec6f> (accessed on 29 October 2023).
114. Lens.org—Detailed Review of a New Open Discovery and Citation Index (No Date) Aaron Tay’s Musings about Librarianship. Available online: <http://musingsaboutlibrarianship.blogspot.com/2018/11/lensorg-detailed-review-of-new-open.html> (accessed on 29 October 2023).
115. Patently transparent. *Nat. Biotechnol.* **2006**, *24*, 474. [CrossRef] [PubMed]
116. Penfold, R. Using the Lens database for staff publications. *J. Med. Libr. Assoc. JMLA* **2020**, *108*, 341. [CrossRef]
117. Graham, S.J.; Marco, A.C.; Miller, R. The USPTO patent examination research dataset: A window on the process of patent examination. In *Georgia Tech Scheller College of Business Research Paper No. WP*; SSRN: Rochester, NY, USA, 2015; Volume 43, Available online: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2729322 (accessed on 29 October 2023).
118. He, Q. Knowledge Discovery through Co-Word Analysis. *Libr. Trends* **1999**, *48*, 133–159.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.