



A closer look on patent analytics through systematic literature review

Suriyati Chuprat¹ · Evan Hanif Dwi Novianto² · Yoshiyuki Matsuura^{2,4} · Akbariah Mohd Mahdzir^{3,4} · Aizul Nahar Harun^{3,4}

Received: 21 July 2023 / Accepted: 10 June 2024

© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2024

Abstract

Patent analytics plays a crucial role in influencing technological innovation and development strategies. However, solely describing the process of patent analytics is insufficient; a critical analysis of the methodologies employed, and their practical implications is required. In this paper, we delve into the realm of patent analytics by conducting a systematic literature review (SLR) of 169 papers from the web of science database. Our analysis, which is driven by six important research questions, aims to provide not only an overview of patent analytics but also a deeper understanding of the topic. In addition to shedding light on the obstacles, tools, techniques, and insights of patent analytics, we also consider uncharted territories and future directions.

Keywords Patent analytics · Technology · Machine learning · Data analytics · Framework · Innovation

JEL Classification O320

✉ Suriyati Chuprat
suriyati.kl@utm.my

¹ Faculty of Artificial Intelligence, Universiti Teknologi Malaysia, Kuala Lumpur, Malaysia

² Graduate School of Innovation and Technology Management, Yamaguchi University Management of Technology, Ube, Yamaguchi, Japan

³ Management of Technology Department, Malaysia-Japan International Institute of Technology, Universiti Teknologi Malaysia, Kuala Lumpur, Malaysia

⁴ International Joint Intellectual Property Laboratory, Yamaguchi University – Malaysia Japan International Institute of Technology, Universiti Teknologi Malaysia, Kuala Lumpur, Malaysia

1 Introduction

In today's fast-paced technological landscape, organizations seeking a competitive edge must secure patents to safeguard their innovations. To devise effective research and development (R&D) strategies, it is essential to comprehend the ever-changing scientific and technological domain and to keep apprised of the most recent developments. Patent analytics, which includes the analysis of scientific publications and patents, has emerged as a valuable resource for businesses. Its applications include business strategy formulation, portfolio analysis, competitor evaluation, technology assessment, and cost analysis.

Patent documents has long been widely recognized by academics and practitioners as a reliable and rigorous source of information for proxy measures of innovation activity (Candelin-Palmqvist et al. 2012; Maragakis et al. 2023; Puccetti et al. 2021; Shen et al. 2020; Shibata et al. 2008). Including technical, market, and legal information, the patent literature is a significant body of work in science and technology (Yanhui and Lixin 2024) that has been assessed as novel and progressive (Adams and Tate 2009; Jang and Yoon 2021). In the continuously changing global market, patent is a valuable company intellectual property, for research and development (R&D) to increase competitiveness by providing technological trends and insight (Jo Kim et al. 2015; Jun et al. 2015; Kim and Bae 2017). Therefore, this research field is one of the critical issues in the forward-looking approaches that have become the focus of innovation management studies in recent years (Naeini et al. 2022). In this study, "patent analytics" is defined as the same as "patent landscaping," which is the act of looking up patents about a particular topic to gauge innovation and evaluate its challenges (Abood and Feltenberger 2018).

Innovation is greatly aided by patent analytics, which provides insightful information extracted from the massive amount of data found in patents. Organizations can obtain a thorough grasp of the state of the art in a particular technology field by examining patent data. This entails tracking patterns in the filing of patents, identifying cutting-edge technologies, and spotting possible holes that could use innovation. Businesses can use patent analytics to inform strategic choices about their R&D (research and development) expenditures. They can learn about the technological orientations of their competitors and pinpoint areas where they can concentrate their R&D efforts to obtain a competitive advantage by looking through competitor patents. Companies can make sure their inventions respect other people's intellectual property rights by examining current patents. This lessens the possibility of pricey patent disputes and legal obstacles. Potential partners or collaborators working on complementary technology can be found through patent analysis. Furthermore, to facilitate access to new markets and technologies, it can identify patents that may be purchased or licensed. Figure 1 depicts an example of an innovation funnel (Ernst et al. 2015). Observed that patent analytics plays an important role in providing insights for all phases in the innovation funnel which include market know-how, technology know-how, technology selection, market acceptance, and strategy for technology concept.

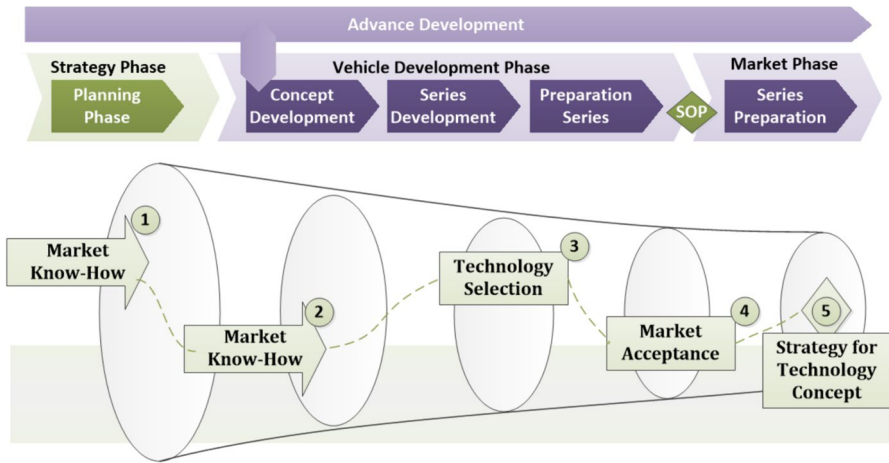


Fig. 1 Example of innovation funnel (Ernst et al. 2015)

There were about 3.4 million published patent applications in 2021, this number is expected to increase over the next 15 years (Adel and Harrison 2024). Furthermore, patent documents encompass a variety of features ranging from bibliographic information (e.g. patent registration office, assignee, inventor, dates, IPC, and citations), textual components (e.g. title, abstract, description, and claims), and image (i.e. drawings) (Trippe 2015). The territorial protection of inventions results in patent families, which consider the geographic jurisdictions of the patent office (Fernandez 2022). Due to differences in patenting authority policies, the three most important dates regarding a patent are the priority date (i.e. filing date at the first patent office), the application date (i.e. filling date), and the publication date (i.e. generally 18 months after the filing date) (Trippe 2015). The classic "innovation funnel" theory has outlined that the creation of successful industrial innovations needs to go through a series of pathways from global to micro levels which transform ideas into inventions that can be commercialized (Dunphy et al. 1996). Hence, these significant increases in volume, various features availability, and volatility of legal context from patent serve wealth references for innovation stakeholders, offering insight for each funnel level spanning from market landscape to technology detail. Nevertheless, in line with the 7 V's¹ of big data definition (Khan et al. 2014), this patent's abundance is a challenge for innovation researchers and managers to view patent analytics no longer just as individual document analysis but as big data analytics (Kim et al. 2021; Park et al. 2018). Therefore, the contemporary patent analytics question is how big data patents can be transformed into business value for companies (Chiu 2018; Jun 2021; Zhang et al. 2019).

¹ Volume, Velocity, Variety, Veracity, Validity, Volatility, and Value which represent characteristics of big data. Each detail in (Khan et al. 2014).

With the digitization of patent data, the largest technical information resource in the world is now available at an affordable price (Aristodemou and Tietze 2018). This is possible because processing ad-hoc data analysis queries over large volumes of patent databases quickly and cheaply can be achieved through cloud services (Antunes et al. 2018). The consolidation of AI, big data, and cloud computing can push the limits of prior expert-dependent patent analysis (Alderucci and Sicker 2019; Lee et al. 2016a, b; Park et al. 2018), by escalating and accelerating strategic technology and innovation decision-making by the innovation manager to be defined objectively from massive patent data (Jun 2021; Lee et al. 2018). Nevertheless, there is an under-adoption risk or even loss of these technological opportunities due to management ignorance influenced by the dependency on traditional best practices (Wang et al. 2016), the uncertainty of the technology's use cases (Ameje et al. 2023) and unhandled information overload (Jackson and Farzaneh 2012; Raguseo 2018). As a result, it is unclear how to divide the white-collar and blue-collar responsibilities and tasks of big data analysis of patent databases for enterprise technology innovation, and it becomes increasingly blurred as the organization scales up, which involves cross-departmental from R&D until Information Technology (IT) teams, to decide strategic innovation and technology infrastructure investment. Therefore, standardization of the patent analytics methodology framework is imperative to develop best practices (Grant et al. 2014).

Previous literature has reviewed a variety of patent techniques and analysis tools which fully or partially facilitated by the AI approach (Abbas et al. 2014; Aristodemou and Tietze 2018; Moehrl et al. 2010) and how to search patent based on the semantic approach and user (Bonino et al. 2010). However, apart from articles on trends of patent databases still understudied (Kim and Lee 2015; Walter et al. 2022a), there is still a lack of specific user targets and its use case in the innovation process (Aristodemou and Tietze 2018), leaving knowledge discrepancy of innovation-interest researchers and users to appropriately determine structured workflow required to conduct patent analytics effectively by optimizing AI, big data, and cloud technology. The era of Artificial Intelligence makes data play a main driver of the innovation management process (Tekic and Füller 2023). In particular, the creation of innovations by companies today highlights how capable they are in handling big data and business analytics (Yoshikuni et al. 2023). The utilization of big data encourages companies to explore hidden corporate values and exploit new opportunities (Raguseo 2018). Furthermore, cloud computing as complementary technology facilitated big data by proceeding online repository system which can reduce costs, improve system performance agility and flexibility, and enable data exchange across corporate boundaries (Bonner et al. 2017; Liu and Xu 2017). Since AI, big data, and cloud computing provide significant cost advantages for companies in data repository and information processing, the current way of innovation needs to be examined (Füller et al. 2022; Haefner et al. 2021; Y. Liu and Mai 2024).

Addressing this requirement, using the Systematic Literature Review (SLR) methodology, this review aims to identify and analyse frameworks, techniques or

algorithms, databases, tools, insight of strategic technology innovation, and big data as well as cloud computing application that used in AI-driven patent analytics research from scientific papers between 2017 and 2022. In contrast to previous studies that classify user-oriented patent analysis tasks (Bonino et al. 2010) and Pat-informatics as ARIS-business process model (Moehrle et al. 2010), in this study, the de facto industry standard for data mining that more suitable for big data era, the CRISP-DM process model (Ncr and Clinton, 1999; Schröer et al. 2021), was adapted to investigate the extent of the standardization deficit for patent analytics frameworks. Given the time frame of this study, our systematic review builds on and expands on previous efforts (Aristodemou and Tietze 2018) that also filled gap with empirical evidence of the patent database usage and the practical know-how of patent analytics when a rapid evolution of AI than its prediction in 2017 (Brynjolfsson and McAfee 2017; Panetta 2018).

Furthermore, this study contributes to the recent progress of patent analytics that has been supported by AI, big data, and cloud technologies in the domain of enterprise technology innovation, especially R&D strategy, and technology analysis process. Set the context from this holistic point of view minimized overlap between knowledge management and technology management of previous patent analysis studies (Aristodemou and Tietze 2018) which provided longitudinal evidence of patent analytics influence on innovation management practices. This study offers managerial implications by compiling use cases that can be used as a guideline for innovation practitioners to exploit and expedite public or commercial patent databases to be valuable technological knowledge to support the innovation management process and offers a coherent future research agenda for researchers to develop more advanced AI, big data, and cloud computing-based patent analytics methodology (Alderucci and Sicker 2019).

This paper is organized as follows. We described the SLR methodology for our research used in Sect. 2. The result and discussion from the selected paper are presented in Sect. 3. We outline our research limitations in Sect. 4. Then, conclusion of this study in Sect. 5.

2 Methodology

This study followed an established process to generate a systematic sample of scientific papers (Kitchenham et al. 2009). The SLR methodology allows to answer a specific research question through identifying all empirical evidence based on predetermined inclusion criteria (Snyder 2019). Bias in findings can be minimized because the selection of studies according to strict and transparent procedures (Snyder 2019). The main objective of the SLR is to provide a comprehensive understanding to innovation-interest researchers and users to appropriately determine the structured workflow required to conduct patent analytics effectively by optimizing AI, big data, and/or, cloud technology. Based on this purpose, clear research questions are needed to define the scope and contribution to academia or practice (Hällgren 2012). It is effective to inquire the

Table 1 Research questions on literature review

ID	Research question
RQ1	Which data science frameworks or processes are commonly used in patent analytics?
RQ2	What techniques or algorithms are prevalent in patent analytics?
RQ3	Which databases serve as primary sources for patent analytics?
RQ4	What tools are frequently employed in patent analytics?
RQ5	What business or strategic technology innovation insights and visualizations are typically produced through patent analytics?
RQ6	To what extent are big data and cloud computing platforms applied in patent analytics?

precise and expressive questions to highlight effective future research work direction. Table 1 shows the research questions that motivated this literature review.

To find primary studies, the following steps must be taken: selecting digital libraries, defining the search string, conducting a pilot search, refining search strings, and retrieving the initial list of primary studies. Sample studies were obtained from scientific publications that had been published in the Web of Science (WOS) database. This database was chosen because it is the oldest, most widely used and authoritative database of research publications and citations in the world (Birkle et al. 2020). A search string was created based on a combination of alternative spellings or synonyms of the term "patent analytics". The following keywords were used in the SLR search:

2.1 "Patent analytics" OR "Patent analysis"

The search for scientific publications was conducted in April 2022. The search was limited to the abstract, title, and keywords of the papers since a large number of relevant papers were published in recent years. From the initial search results, the filtering feature on the WoS website was used to efficiently identify papers that were relevant to research interests. As an update to the study from (Aristodemou and Tietze 2018) and examine the impact of developments in artificial intelligence that have been heralded since 2017 (Brynjolfsson and McAfee 2017; Panetta 2018) specifically on patent analytics studies, we set the publication year 2017–2022. We set journal articles, review articles, proceeding papers, and book chapters as sample inclusion criteria to capture theoretical and practical works of literature. To avoid misinterpretation, papers that do not use English are excluded from the samples. Figure 2 outlines the selection process of papers for the systematic literature review using the PRISMA method (Moher et al. 2009).

After supported filtering, paper selection was carried out based on the study content. One of the authors analyzed this list and excluded articles that were only partially related to our research and/or outside the scope of this special issue, which focuses on patent analysis for strategic technology and innovation leveraging AI technologies, big data, and/or cloud computing. The selection process is divided into two stages. The first stage is screening through the title and abstract. By perform quality checking, the second stage is an in-depth study by reading the full text to ensure eligibility. Papers that pass these two stages become the final samples for the systematic review and further

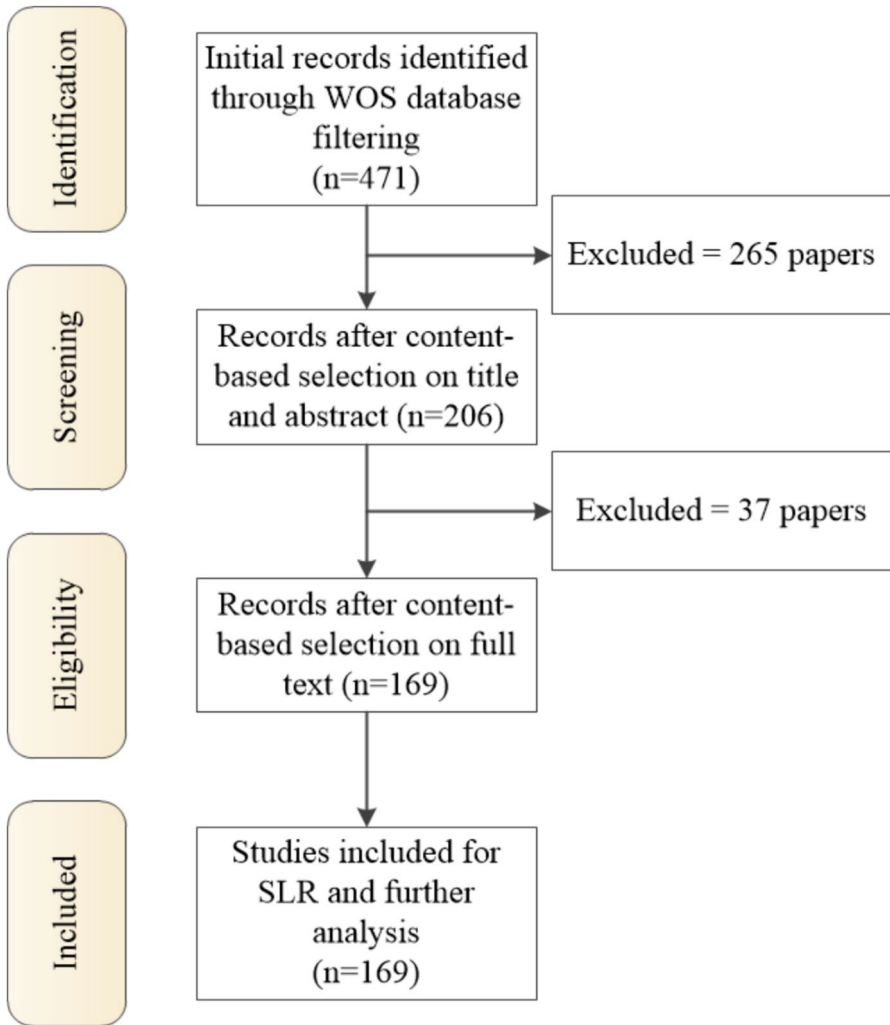
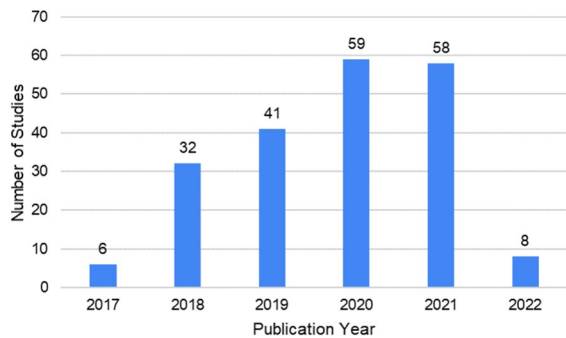


Fig. 2 PRISMA flow diagram for SLR

analysis. Then, we meticulously extracted data from the selected papers, organizing it into distinct categories detailed in Table 2. The organization of the paper by the authors in this extraction format allows for concept-centric synthesis (Kraus et al. 2020). This extracted data was subsequently catalogued and analyzed using Microsoft Excel and Nvivo.

Table 2 Extraction form

Field	Extracted data
1	Paper_ID
2	Author
3	Year
4	Framework/methodology/process
5	Techniques/algorithms
6	Database
7	Tools
8	Insights
9	Big data/cloud

Fig. 3 Patent analytics publication trends

3 SLR analysis and discussion

This section presents the analysis results of the SLR based on the six research questions outlined in the previous section. The analysis was done on the 169 selected papers using MS-Excel and NVivo. Figure 3 shows the number of papers according to year and it depicts the growing trend on patent analytics since 2017. Note that some of the papers published in 2017 were included, but it was registered online in 2018.

3.1 RQ1-data science framework and process used for patent analytics

The cross industry standard process for data mining (CRISP-DM) is a process model that serves as the base for a data science process. As shown in Fig. 4, CRISP-DM was published in 1999 to standardize data mining processes across industries, since then has become the most common methodology for data mining, analytics, and data science projects. Generally, it has six sequential phases as follow:

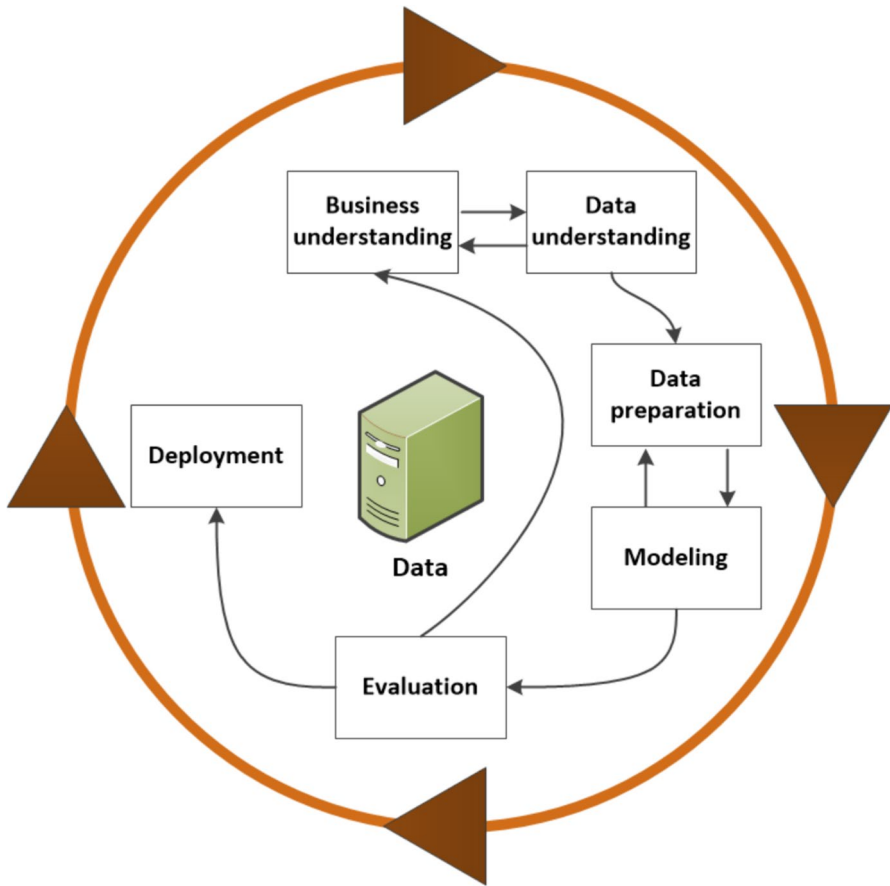


Fig. 4 Cross industry standard process for data mining (IBM 2021)

- i. **Business understanding:** understanding the needs and commercial goals for the data mining project are the main goals of this phase. Determining the target audience, outlining the issue statement, and establishing the project's success criteria are all included in this.
- ii. **Data understanding:** understanding the data that will be used in the project is the main goal of this phase. This entails examining the data, spotting problems with its quality, and recording its features.
- iii. **Data preparation:** data preparation for modelling is the main goal of this phase. Cleaning, converting, and choosing the pertinent features for modelling are all included in this process.
- iv. **Modelling:** data mining model construction and training are the main goals of this phase. This encompasses the process of choosing suitable modelling methodologies, developing the models, and assessing their efficacy.

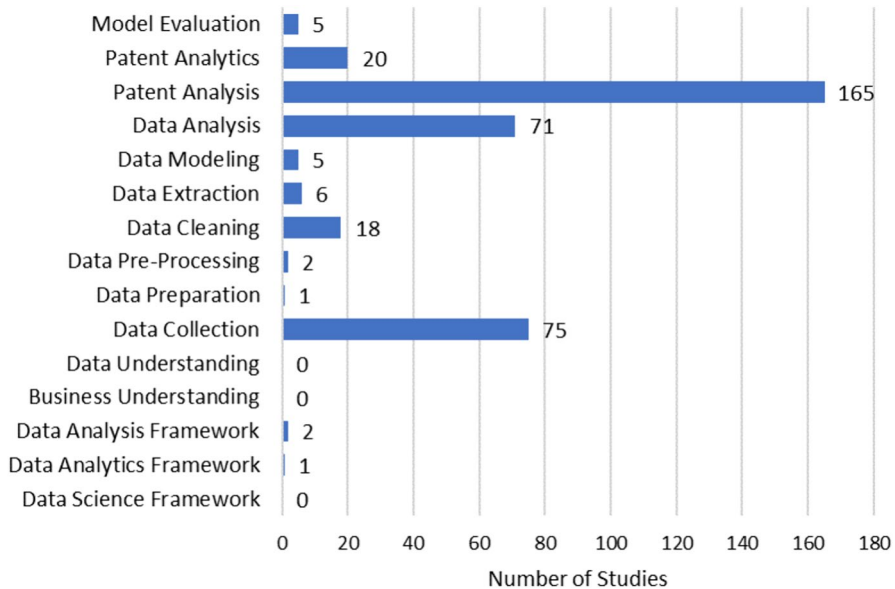


Fig. 5 RQ1 patent analytics process

- v. Evaluation: this stage is devoted to assessing the models' performance and choosing the most suitable model for implementation. In order to do this, the models' memory, accuracy, and precision are evaluated and contrasted with the business goals.
- vi. Deployment: putting the chosen model into production is the main goal of this phase. This entails monitoring the model's performance, updating it as necessary, and integrating it into an already-existing system.

We extracted related patent analytics processes using the CRISP-DM as a guideline, and the analysis results are shown in Fig. 5. We also used terms that are similar to those used in CRISP-DM, as some of the terms are sub-tasks in CRISP-DM. For example, data collection is a component of data understanding. Our research reveals a startling omission in the landscape of patent analytics: the absence of emphasis on “*Business Understanding*” and “*Data Understanding*”. Despite the fact that these terms are essential to CRISP-DM, their absence from the discourse of selected papers is striking. In the context of patent analytics, understanding specific business or strategic goals is not merely advantageous; it is essential. Without a clear understanding of the organization's objectives and success criteria, any analytics project runs the risk of being ineffective.

In addition, “*Data Understanding*” is similarly underrepresented. Understanding the nature, quality, and limitations of the data at hand is required for meaningful analysis in the era of big data. Ignoring this phase can result in unbalanced outcomes and lost opportunities. The underemphasis on preparatory phases such as data collection, preprocessing, and cleansing is also noteworthy. These are the

ungung champions of data analytics, the processes that transform raw data into a usable format. In patent analytics, where data can be especially unstructured and chaotic, ignoring these preparatory phases can have disastrous effects on the caliber of insights obtained. What are the implications of these observations for the real world? First, practitioners of patent analytics must recognize that CRISP-DM is not a rigid doctrine. It should be modified to accommodate the specific requirements of patent-related initiatives, with a focus on “Business Understanding” and “Data Understanding”. In addition, there should be a renewed emphasis on the preparatory phases, recognizing their importance in determining the success of a project. In conclusion, this transition from description to critical analysis highlights the significance of combining well-established frameworks, such as CRISP-DM, with the particular demands of patent analytics. By doing so, we can unlock the full potential of data science to fuel innovation and strategic decision-making in the domain of intellectual property. Patent analytics can evolve into a more effective and strategic tool for organizations and researchers equally by evaluating the approaches critically and considering their practical implications.

3.2 RQ2-algorithms and techniques used for patent analytics

Patent-based measurements are quantitative indicators obtained by studying patent data. They are useful tools to evaluate many elements of innovation and technology. Measures include technological novelty, which seeks to capture a patent’s newness and inventiveness by assessing characteristics such as citations, technical keywords, and claim structure. Technological impact is a measure that analyzes the importance and relevance of a patent by examining forward citations, which are citations obtained from other patents. Innovation activity is the measure that tracks the overall innovation output and trends within a certain area or organization by tracking patents, patent applications, or inventor activity. For these measures to be calculated and analyzed, many techniques from patent analytics could be used. Table 3 shows example of extracted data from patent files, calculated measures and used techniques.

The realm of patent analytics is far from monolithic, characterized by a diverse toolbox of techniques and algorithms aimed at extracting meaningful insights from patent data. While K-Means Clustering and Latent Dirichlet Allocation (LDA) dominate the field, a comprehensive view necessitates a deeper understanding of the multifaceted approaches employed. We observed that most papers mentioned the use

Table 3 Common patent-based measurements

Calculated measures	Extracted data	Patent analytics techniques
Technological novelty	Citation, technical keywords, claim	LDA (topic modelling) Text mining techniques
Technological impact	Forward citation	Citation analysis
Innovation activity	Patent application	Assignee grammar analysis Machine learning algorithms

of data mining, machine learning and deep learning. Nevertheless, we know that text mining is an essential process of exploring and analyzing large amounts of unstructured text data using software that can identify concepts, patterns, topics, keywords, and other attributes in the data. Text mining is the basis of patent analytics, enabling us to deal with enormous quantities of unstructured textual data. An interpretive lens reveals the crucial function of natural language processing (NLP), a powerful tool that simulates human text comprehension. NLP combines computational linguistics with statistical, machine learning, and deep learning models to enable the extraction of nuanced insights. The incorporation of NLP into patent analytics not only improves the efficiency of data processing, but also reveals latent patterns and connections that would otherwise remain hidden using conventional methods. This in-depth examination of textual data increases the potential for innovation by revealing concealed opportunities and threats.

In structuring patent data, clustering techniques such as Self-Organizing Map (SOM) and the Louvain algorithm play a crucial role. These algorithms provide a nuanced view of technological trends and competitive landscapes by categorizing patents according to their inherent similarities. It is crucial to comprehend the strengths and limitations of clustering algorithms. Effective application of the SOM or Louvain algorithm can enlighten strategic decisions, such as portfolio management or the identification of prospective partners. However, improper application can result in erroneous conclusions, highlighting the significance of algorithm selection and parameter tuning.

Association Rule Mining (ARM) arises as a potent technique for patent analytics, revealing complex relationships between patents. This technique surpasses basic clustering by disclosing hidden patterns of co-occurrence and interdependence among patent filings. The strategic value of ARM's ability to unearth concealed associations is significant. It can aid patent examiners in more effectively identifying prior art and help companies identify potential infringement risks. To distinguish between false associations and meaningful patterns, however, a nuanced approach is required. Word Embedding and Multi-Layer Perceptron (MLP) are gaining traction as the field of patent analytics advances. Word Embedding facilitates a deeper comprehension of the semantic relationships between terms, whereas MLP can model complex, nonlinear interactions in patent data. Exploration of these emerging techniques can provide a competitive advantage. Word Embedding can enhance patent search and classification, thereby improving the accuracy of patents that are retrieved. MLP, on the other hand, enables more sophisticated patent valuation and forecasting predictive models. The remaining analytics techniques and algorithms used in patent analytics are as listed in Table 4.

A critical examination of patent analytics techniques and algorithms exposes a dynamic landscape with profound practical implications. Integration of NLP, comprehension of clustering nuances, exploitation of ARM's potential, and exploration of emerging methods are crucial next stages. Patent analytics can become a transformative instrument for innovation, intellectual property strategy, and informed decision-making in an ever-changing technological landscape by navigating this complex terrain.

Table 4 RQ2 patent analytics techniques

Analytics techniques	Type	References
Latent Dirichlet allocation (LDA)	AI-based algorithm	Ahn et al. (2021), Altuntas and Aba (2022), Chao et al. (2021), Choi and Song (2018), Chung et al. (2021), Feng et al. (2021), Feng et al. (2020), Garzanti et al. (2021), Geum and Kim (2020), Jain et al. (2020), Kang et al. (2019), Kang et al. (2021), Kim et al. (2019a, b), Kim et al. (2019a, b), Lee et al. (2021), Liu et al. (2022), Song and Suh (2019), Trappey et al. (2020), Xu and Gui (n.d.), Yu et al. (2020)
Partitioning around medoids (PAM)	Rule-based techniques	Jain et al. (2020), Yu et al. (2020)
Dimensionality reduction methods	Feature selection technique	Krestel et al. (2021)
Rapid automatic keyword extraction (RAKE) algorithm	NLP	Du et al. (2021)
Fruchterman reingold algorithm	Rule-based techniques	de Paulo et al. (2018)
MST algorithm	Optimization algorithms	Comai (2018)
Self-organizing map (SOM) algorithm	AI-based algorithm	Altuntas et al. (2020), Cho et al. (2021), Liu et al. (2020), Yu et al. (2020)
A priori algorithm	Rule-based techniques	Feng et al. (2020)
Concor algorithm	Rule-based techniques	Liu et al. (2018), Yuan and Li (2021)
Support vector regression (SVR)	AI-based algorithm	Lee et al. (2016a, b)
Louvain algorithm	Rule-based techniques	Aaldering et al. (2019), Chung et al. (2021), Minghan et al. (2020), Zhou et al. (2021)
Cluster proximity index measurement	Rule-based techniques	Zhou et al. (2021)
Fuzzy clustering algorithms	AI-based algorithm	Arnaldo Valdés et al. (2019)
K-means	AI-based techniques	Altuntas et al. (2020), Ampornphan and Tongngam (2020), Chao et al. (2021), L. Feng et al. (2020), Guiderian (2019), Jain et al. (2020), Jang et al. (2021), Kang et al. (2019), Kim et al. (2019a, b), Krestel et al. (2021), M. Lee and He (2021), S. Lee et al. (2022), Y. Liu et al. (2022), Mao et al. (2022), Minghan et al. (2020), Sasaki and Sakata (2021), Smojver et al. (2019), Thavorn et al. (2021), Yu et al. (2020), Zhou et al. (2018)
Genetic knowledge persistence algorithm	Rule-based TECHNIQUES & OPTIMIZATION	Feng and Magee (2020)
Newman clustering algorithm	Rule-based techniques & AI-based Algorithm	Zhou et al. (2018)

Table 4 (continued)

Analytics techniques	Type	References
Association rule mining (ARM)	Rule-based techniques & AI-based algorithm	Amornphan and Tongngam (2020), Cho et al. (2021), Park et al. (2021), Thavorn et al. (2021)
GMM clustering algorithm	AI-based algorithm	Lee et al. (2022)
K-Medoids	Rule-based techniques	Jain et al. (2020)
Support vector machines (SVM)	AI-based algorithm	Liu et al. (2020)
Multi-layer perceptron (MLP)	AI-based algorithm	Chiarello et al. (2018)
Word embedding algorithm	NLP	Chen et al. (2020)

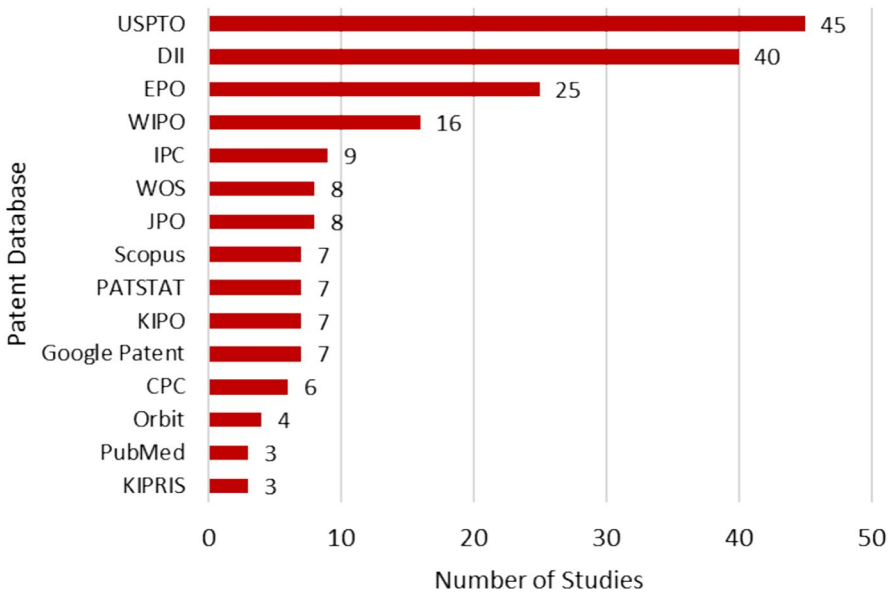


Fig. 6 RQ3 most used databases in patent analytics

3.3 RQ3-databases used in patent analytics

Patent analytics is based on databases, which provide access to an extensive repository of intellectual property data. However, a simple count of their popularity is insufficient for elucidating their significant impact. We delve deeper to comprehend not only which databases are frequently utilized but also why they matter in the ecosystem of patent analytics. Figure 6 depicts the most used database in patent analytics. Based on our SLR analysis, USPTO was ranked highest, followed by DII/DWPI, EPO and WIPO. We only included databases which had at least 3 frequencies to be displayed in Fig. 6. Other databases used in patent analytics are CNIPA, ESPACENET, INNOGRAPHY, OECD, SIPO, Acclaim, CCMT Database, CITES, CrunchBase, CureVac, DPMA, EC, FIPS, FoundationIP, GPI, Inteum, IPD, IPTECH, Ipzen, IWT, Nanosats Database, NewSpace Hub, PatBase, Patentinspiration, Patseer, Questel-Orbit, SciVal, Symphony Innovate, ThomsonIP Manager, WO-PCT and PCT.

The United States Patent and Trademark Office (USPTO) is the national patent office and trademark registration authority for the United States, and it is part of the Department of Commerce. According to the most recent report issued by the USPTO in May 2021, a total of 388,900 patents were granted in 2020—underscores its significance. It was also reported that over 7.8 million US patents have been issued since the first was granted on July 31, 1790, including over 7.2 million utility patents ("patents for inventions"). To make use of this vast body of technology easier, these documents have been "classified" (categorized) into roughly 470 broad technological categories (called classes) and approximately 159,000 specific

technological categories (called subclasses) that comprise the United States Patent Classification System (<https://www.uspto.gov/>). The USPTO's role in patent analytics extends beyond borders. Its extensive collection empowers researchers worldwide to access, analyze, and draw insights from American innovation. For businesses, it serves as a valuable resource for competitive intelligence and innovation strategy.

Derwent Innovation, specifically the Derwent World Patents Index (DWPI), offers more than just a repository of patents. Derwent Innovation has two main features: the Derwent World Patents Index (DWPI) and the smart search capability. After specialists have studied the whole official patent disclosure materials, DWPI perform the translation, rewrite the important abstracts, content debugging, and normalization of patent holders. Novelty, use, benefit, technical focus, full description, drawing description, activity, and mechanism are among the DWPI revised components (Walter et al. 2022b). It is also claimed that DWPI improved keyword search results by 79% when compared to patent searches performed on other patent search platforms without DWPI (<https://clarivate.com/derwent/>). This could be one of the reasons why DWPI is one of the top five databases used in patent analytics. The data refinement provided by DWPI transforms patent analytics into a precise instrument. Researchers can effectively navigate the complexities of patent jargon, while businesses gain a competitive advantage in discovering pertinent prior art and identifying emerging trends.

The European Patent Office (EPO) examines European patent applications, allowing inventors, researchers, and businesses from all over the world to obtain protection for their inventions in up to 44 countries via a centralized and uniform procedure requiring only one application. According to a report issued in May 2022, the EPO has published over four million patent applications (<https://epo.org>). With over four million patent applications published, it promotes a unified procedure, thereby facilitating the acquisition of European patents. EPO provides a streamlined entry point to European markets for businesses pursuing international expansion. It simplifies the complex process of obtaining patent protection in multiple jurisdictions, thereby facilitating effective portfolio management.

The World Intellectual Property Organization (WIPO) boasts a unique global reach, serving 193 member nations. It is dedicated to intellectual property (IP) services, policy, information, and collaboration. The PATENTSCOPE database of WIPO provides access to international Patent Cooperation Treaty (PCT) applications in full text format on the day of publication, as well as to patent documents of participating national and regional patent offices. There are more than 104 million patent documents, including international patent applications submitted under the PCT (<https://www.wipo.int/patentscope/en/>). The WIPO database exemplifies global cooperation in patent analytics. It enables researchers and businesses to navigate a maze of international patent data, fostering innovation and facilitating global decision-making.

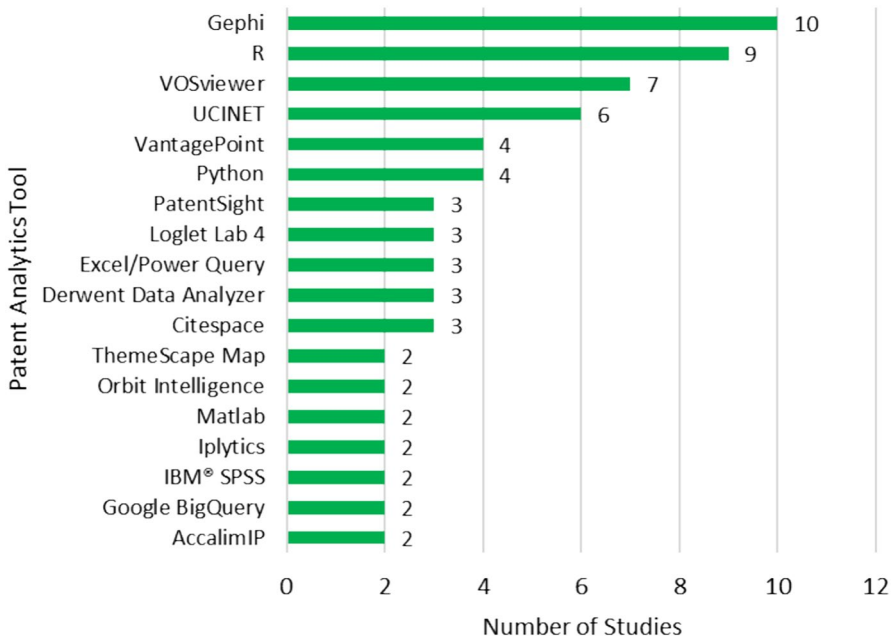


Fig. 7 RQ4 most used tools in patent analytics

3.4 RQ4-tools used for patent analytics

This section guides us in understanding not just the tools frequently employed but also how they shape patent analysis, innovation, and intellectual property strategies. Tools are the core component of patent analytics, bridging the chasm between data deluge and actionable insights. However, an in-depth comprehension of the tools necessitates a critical evaluation of their impact on patent analysis and its practical implications, in addition to a listing of their popularity. Figure 7 shows the most used tools in patent analytics. Our study reveals that although commercial software dominates (e.g. PatentSight, VantagePoint, UCINET, etc.), The top three widely used tools in patent analytics studies are open-source software. Other tools which are not displayed in the graph of Fig. 7 are Bibliometrix, Chinese Knowledge Information Processing (CKIP), CiteNet, Core NLP, Cytoscape, Derwent Smart Search (SSTO), Doc2vec, EconSight, Essential Science Indicators, ggplot2, Hive, IPTECH software, Leximancer, Mathematica, NetMiner, NVivo, Open NLP, Patent iNSIGHT Pro, PatentsView, spaCy, Tableau, TextBlob, Vensim@ PLE, Weka and WordStat.

Gephi is the industry standard for graph and network visualization and exploration software (Bastian et al., n.d.). Gephi was chosen as a tool for building, visualizing, and analyzing patent networks because it is open-source software with a user-friendly interface that allows for agile and accurate analyses based on Social Network Analysis (SNA) theory (de Paulo et al. 2018). Miao et al. (2022) used Gephi to build a multilayer network of Technology Relationship Technology (TRT) and then thoroughly analyzed the structural relationships between the dimensions

to extract the composition, functions, means, space, time, and advantages of technical components. Gephi was chosen by Zhang et al. (2021) for its ability to calculate and visualize large and complex networks. The network analysis prowess of Gephi enhances the capacity to extract vital insights from patent data. Researchers and innovators can decipher complex relationships, identify emergent trends, and make decisions based on data to promote innovation and competitiveness.

R is a statistical computing and graphics language and environment. It is highly extensible and offers a wide range of statistical (linear and nonlinear modelling, classical statistical tests, time-series analysis, classification, and clustering) and graphical techniques. R is also available as Free Software under the Free Software Foundation's GNU General Public License in source code form, which can be compiled and run on a wide range of UNIX platforms and similar systems (including FreeBSD and Linux), Windows, and MacOS. Yu et al. (2020) performed preprocessing through R programming which include tokenizing and removing stop-words. These processes were conducted to increase the data reliability. Yalcin and Daim (2021) used R programming for cleaning, compiling, and analysis of data.

VOSviewer is a tool for creating and visualizing bibliometric networks. These networks can be built using citation, bibliographic coupling, co-citation, or co-authorship relationships, and can include journals, researchers, or individual publications. VOSviewer also includes text mining functionality for creating and visualizing co-occurrence networks of key terms extracted from scientific literature. Dehghani et al. (2021) used VOSviewer software to construct and visualize patents keyword networks. With aid of domain experts, LDA and the VOSviewer mining software, Feng et al. (2021) performed data preprocessing to extract 26 topics with few other keywords from 255 documents. The text mining functionality of VOSviewer equips patent analysts with a potent instrument for discovering hidden patterns in patent data. This results in more accurate technology landscapes, which aids in identifying strategic directions and innovation opportunities.

UCINET is a comprehensive package for the analysis of social network data. It can read and write a multitude of differently formatted text files, as well as Excel files. Shi and Zhang used UCINET for their data processing (Shi and Zhang 2018), while Yuan and Li (2021) used UCINET to partition countries by the CONCOR algorithm. On the other hand, UCINET can be used for visualization (Nordensvard et al. 2018). The ability of UCINET to read and write multiple data formats empowers researchers at multiple phases of the analysis procedure. UCINET enables a comprehensive perspective on patent networks, enabling the identification of key actors, communities, and structural insights. This comprehension guides strategic partnerships, analysis of the competitive landscape, and innovation strategies.

Another popular tool is VantagePoint (VP), a professional-grade desktop text mining application offering analysts a broad suite of powerful refining, analyzing, and reporting tools for scientific, technical, market and patent information. Alvarez-Meaza et al. (2020) utilized VP for data cleaning, while VP also can be used for

visualization (Zhao 2018). The capabilities of VP refine patent information, transforming it into a dependable source of scientific, technical, market, and patent intelligence. The VP's role in data cleansing ensures that patent analytics are founded on a solid foundation, thereby reducing the number of inaccurate insights. Researchers and analysts can have confidence in the veracity of their findings, which facilitates more informed innovation and IP strategies.

In conclusion, patent analytics tools are catalysts for innovation and strategic decision-making rather than mere utilities. Each tool contributes a distinct set of capabilities that shape the insights and pathways derived from patent data. However, the selection of tools is a complex decision that is influenced by data type, analysis requirements, and desired outcomes. Recognizing the interplay between these tools and other criteria is essential for maximizing the potential of patent analytics in a technological landscape that is swiftly evolving. By critically scrutinizing their roles, we shed light on the path to more efficient and consequential patent analysis practices.

3.5 RQ5-visualization and insights in patent analytics

In the domain of patent analytics, innovators, businesses, and strategists seek insights as prized assets. However, the true essence of these insights goes beyond enumeration; it beckons us to discover their transformative effect on innovation, competition, and intellectual property strategies. Figure 8 depicts the major insights generated by the patent analytics. According to our analysis, descriptive insights serve as the cornerstone of patent analytics. Among these, insights on patents developed by country and assignee occupy the center of research papers and practical implementations. The prevalence of descriptive insights demonstrates their usefulness. Businesses can use this data to assess the global innovation landscape, identify

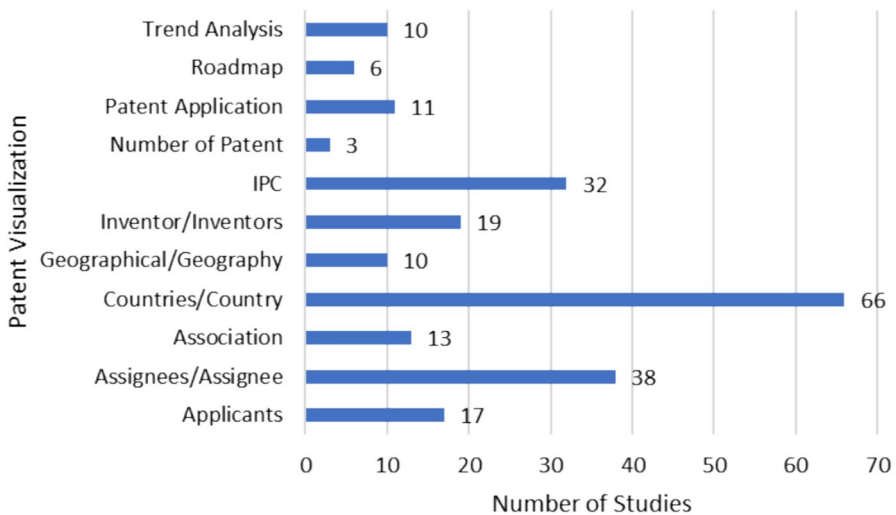


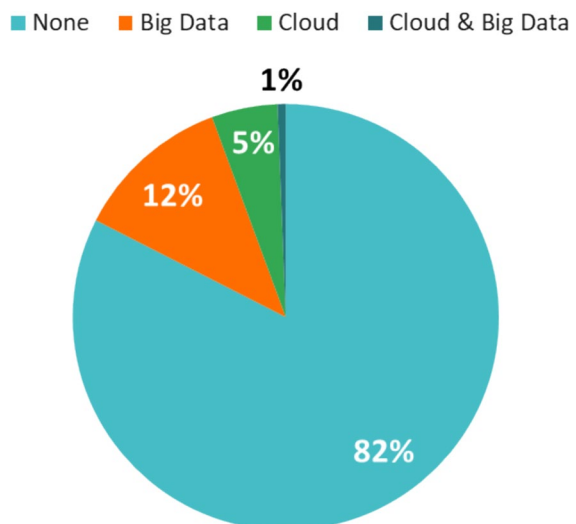
Fig. 8 RQ5 Insights in patent analytics

potential collaborators or competitors, and make informed decisions regarding geographic expansion or partnership opportunities.

In addition to descriptive insights, patent analytics has evolved to uncover deeper insights, such as technological trends. These insights delve deep into innovation, providing a nuanced comprehension of emerging technologies and their trajectories. Understanding technological trends equips organisations with the foresight necessary to align their strategies with emerging markets and technologies. This enables them to remain ahead of the curve, make wise investments, and seize opportunities in industries that are rapidly evolving. Patent analytics now includes analysis of the competitive landscape, a crucial aspect for businesses. This involves evaluating the patent portfolios of competitors, identifying vacant spaces for innovation, and estimating the likelihood of patent infringement. Businesses attempting to maintain a competitive advantage must conduct a competitive landscape analysis. It guides IP strategies, assisting businesses in protecting their innovations and averting legal pitfalls, thereby securing their market positions. The mapping of the innovation ecosystem is an expansion of patent analytics. This involves identifying key actors, innovators, and influential entities in specific technological domains, thereby shedding light on potential collaboration or acquisition opportunities. The innovation ecosystem can be mapped for strategic advantage. The data-driven insights derived from patent analytics can assist businesses in fostering collaborations, forming partnerships, and exploring acquisition opportunities.

In conclusion, patent analytics yield insights that transcend mere data elements. Businesses and strategists use them as a compass to navigate the complex terrain of innovation and intellectual property. A foundation of descriptive insights provides a panoramic view of patent landscapes. However, it is the deeper insights—technological trends, competitive analysis, and mapping innovation ecosystems—that provide the transformative power to shape innovation strategies, enhance competitiveness,

Fig. 9 RQ6 Application of big data and cloud computing in patent analytics



and drive success. We reveal the true potential of patent analytics as a catalyst for informed decision-making, innovation, and intellectual property strategies in an ever-changing technological landscape by critically evaluating the nature and practical implications of these insights.

3.6 RQ6-big data and cloud used in patent analytics

Patent data is uniquely suited for big data tools and techniques, because of the high volume, high variety (including related information) and high velocity of changes. However, surprisingly, we discovered through the analysis of RQ6, we observed that, not many papers clearly mentioned the used of big data tools or cloud platform. As shown in Fig. 9, only 18% specifically mentioned the used of it. For businesses and researchers, the underrepresentation of big data and cloud computing in patent analytics is a missed opportunity. These technologies have the potential to revolutionize patent data analysis by facilitating real-time processing, scalability, and deeper insights. The limited mention emphasizes the need for increased awareness and implementation.

The velocity of patent analytics is revolutionized by big data and cloud computation. They facilitate real-time data processing, ensuring that businesses and researchers have access to the most recent data, which is essential for making prompt decisions. Patent analytics are significantly affected by real-time insights. Businesses can monitor competitive landscapes, identify emergent trends, and respond rapidly to patent landscape shifts. This adaptability can be a game-changer in innovation and intellectual property strategy. The volume and diversity of patent data necessitate the inherent scalability of cloud computing platforms. This facilitates the exponential development of patent applications by enabling patent analysts to manage enormous datasets. Scalability ensures that patent analytics can keep up with the continually expanding volume of patent data. This is crucial for organizations seeking to maintain complete and accurate patent landscapes, which facilitates strategic decision making and innovation.

Cloud computing reduces costs by eliminating the need for substantial infrastructure investments. Researchers and businesses have on-demand access to powerful computational resources, paying only for what they consume. Patent analytics become accessible to a broader spectrum of organizations as a result of cost-effectiveness. Smaller businesses and entrepreneurs can utilize patent analytics without incurring significant up-front costs. In patent analytics, the potential of big data and cloud computing is mainly uncharted. Although these technologies promise real-time insights, scalability, and cost-effectiveness, their limited mention in research and practice demonstrates the need for greater awareness and adoption. By recognizing the transformative power of these tools, patent analytics can become a more agile, comprehensive, and accessible discipline, empowering businesses and researchers to navigate the complex intellectual property and innovation landscape of the twenty-first century.

4 Limitation

The systematic literature review (SLR) conducted between 2018 and 2022 sheds light on the transforming patent analytics landscape. However, as we interpret the findings, it is essential to recognize the limitations imposed by our research's scope and methodology. Our findings are derived from a subset of the available patent analytics literature due to our reliance on the Web of Science database contained within UTM's subscriptions. To encompass the entirety of this dynamic field, a broader investigation is required, as innovations may occur outside of this confined area.

The 169 selected papers in our SLR constitute a diverse tapestry of research on patent analytics. Nonetheless, while the quantity is substantial, the crucial issue is whether this body of work has explored the field's complexities and depths. The overwhelming number of chosen papers is evidence of the expanding interest in patent analytics. It also encourages researchers and practitioners to delve deeper, nurturing a culture of inquiry that seeks to unearth hidden insights and propel innovation in the field.

Our research scope, which spans 2018 to 2022, provides a temporal context for the evolution of patent analytics research. It provides a summary of the state of the art, encompassing the most recent developments, methodologies, and applications in the field. The context of time emphasizes the dynamic character of patent analytics. Businesses and researchers must acknowledge that the field is in a constant state of change, necessitating adaptability and a forward-thinking approach to maximize its potential.

While our SLR provides vital insights, it is inherently limited. The limited focus on a particular database and subscription journals may omit innovative work published in other repositories and outlets. This restriction necessitates a more expansive approach to patent analytics research. It is essential to investigate a variety of sources and platforms in order to comprehend the breadth of insights and methodologies influencing the field. Our SLR summarizes both the vitality and limitations of patent analytics research. By acknowledging the limits of our study, we open the door to broader horizons and unrealized possibilities. We chart a course for the field to evolve, adapt, and drive intellectual property strategies and innovation in the ever-changing landscape of the twenty-first century by embracing these limitations as innovation catalysts.

5 Conclusion and future work

This article significantly contributes to the development of patent analytics research. Through a systematic literature review (SLR) of 169 articles published between 2018 and 2022, we intend to redefine the direction of the discipline. Our investigation into the methodologies employed in patent analytics has yielded crucial insights, resulting in a deeper comprehension of its practice and practical implications. As we reach the conclusion of our SLR, we find ourselves at the intersection of description and

interpretation, where the richness of patent analytics becomes apparent. Our primary objective was to discover the essence of patent analytics, and our journey has revealed both clarity and complexity. The lack of a standardized data science framework emphasizes the inherent diversity and adaptability of patent analytics. This realization compels us to consider a more adaptable strategy, one that can accommodate the specific requirements of innovation and intellectual property strategies.

While a standard framework for data science remains elusive, our analysis reveals the essence of patent analytics. The majority of papers presented their own process flow or framework, with a focus on the fundamental data science processes of data acquisition, data cleaning, data analysis, and data visualization. This decentralization suggests that patent analytics is a dynamic field that is continuously evolving to meet specific challenges and needs. Patent analytics are dynamic, as evidenced by the prevalence of customized frameworks. Businesses and researchers should embrace this adaptability by customizing their approaches to meet their objectives and effectively navigating the complexity of patent data.

Our SLR provides a panoramic view of the landscape of patent analytics, illuminating the most popular databases, analysis techniques, and tools. We summarized all our SLR findings with the alignment towards CRISP-DM framework as shown in Table 5 in Appendix. This comprehensive comprehension equips businesses and researchers with the skills necessary to effectively leverage the power of patent data. The insights into frequently utilized resources enable stakeholders to make informed decisions regarding database subscriptions, technology adoption, and tool selection. This choice can have a substantial effect on the quality and breadth of patent analytics. The scant mention of big data and cloud platforms in patent analytics is a startling discovery uncovered by our research. This absence raises concerns about unrealized potential and the need for a broader understanding of these transformative technologies. This distinction functions as a call to action. Businesses and researchers should investigate the potential of big data and cloud computing to improve patent analytics in terms of real-time insights, scalability, and cost-effectiveness.

In conclusion, this SLR bridges the gap between patent analytics descriptions and crucial insights. Our journey reveals a field that challenges a one-size-fits-all framework, emphasizing its adaptable and dynamic nature. From adopting customized approaches to leveraging the power of big data and the cloud, the implications are vast. As we move forward with these findings in hand, our vision transcends the confines of this investigation. We aim to devise a comprehensive framework and set of guidelines for patent analytics, not just for innovators but for all stakeholders. This framework will serve as a beacon, directing the field towards deeper insights, more informed strategies, and greater innovation in the constantly evolving intellectual property and technology landscape.

Appendix

See Table 5.

Table 5 Summary of patent analytics from SLR

CRISP-DM phases (RQ1)	Quick references based on SLR findings
Business understanding	<p>Example of patent analytics project objectives</p> <p>Technology forecasting/prediction</p> <p>Identify emerging technology</p> <p>Technology/development trends</p> <p>Technology proximity</p> <p>Technological novelty</p> <p>Technological impact</p> <p>Innovation activity</p> <p>Quality classification</p> <p>Technology opportunity</p> <p>Patent quality/valuation</p>
Data understanding	<p>Available databases: (RQ3)</p> <p>USTPO/DI/DWPI/EPO/WIPO/CNIPA/ESPACENET/INNOGRAPHY/IO ECD, SIPO/Claim/IP/CCMT Database/CITES/GrunchBase/CureVac/DP MAIEC/IPSIFoundation/IP/GPI/Inteum, IP/DI/TECH/Ipzen/IW/TINanosats Database/NewSpace Hub/PatBase, Patentinspiration/Patseer/Questel- OrbitSci/VallSymphony Innovate/ThomsonIP Manager/IWO-PCT/ICTI</p>
Data preparation	<p>Example of patent data fields</p> <p>Title, Application Number, Inventor, Applicant, Assignee, Inventor, Country, Publication Date, Year, IPC, CPC, Abstract, Claim, Citations</p>
Modelling	<p>Example of techniques used: (RQ2)</p> <p>K-Means</p> <p>Partitioning around medoids (PAM)</p> <p>Dimensionality reduction methods</p> <p>Rapid automatic keyword extraction (RAKE) algorithm</p> <p>Fruchterman reingold algorithm</p> <p>Louvain algorithm</p> <p>Cluster proximity index measurement</p> <p>Fuzzy clustering algorithms</p> <p>Genetic knowledge persistence algorithm</p> <p>Newman clustering algorithm</p>

Table 5 (continued)

CRISP-DM phases (RQ1)	Quick references based on SLR findings	
Evaluation	MST algorithm	Association rule mining (ARM)
	Self-organizing map (SOM) algorithm	GMM clustering algorithm
	Latent Dirichlet allocation (LDA)	K-Medoids
	Apriori algorithm	Support vector machines (SVM)
	Concor algorithm	Multi-layer perceptron (MLP)
	Support vector regression (SVR)	Word embedding algorithm
	Example of patent insights: (RQ5)	Example of tools used for analysis and visualization: (RQ4)
	Patent trends (year, country, applicant, assignee, counts)	AccalimIPR Loglet Lab 4 Citespace
	Patent relationship	Gephi VOSviewer Derwent Data Analyzer
	Patent strength	Doc2vec Patent Sight Python Excel
	Technology priority ranking	Google BigQuery IBM@ SPSS Iplytics
	Patent portfolio	Leximancer Mathematical Matlab
	Competitive benchmarking	Vensim@ PLE NetMiner Open NLP
Technology road-mapping	Core NLP Orbit Intelligence	
IPC classification	Cytoscape Patent iNSIGHT Pro3	
Patent technology-function matrix	PatentScope PatentsView	
Global distribution of major countries/regions	PatVisorVR Tableau Power Query	
Top organizations worldwide	ThemeScope Map UCINET VantagePoint	
Top inventor analysis		
Possible deployment platform: (RQ6)		
Cloud computing platform & big data		
Deployment		

Acknowledgements We would like to express our sincere gratitude to the reviewers for their invaluable feedback and insightful comments, which have significantly enhanced the quality of this work. Their thorough and constructive reviews have provided us with important guidance, enabling us to refine our research and present it more clearly.

Author contributions Suriyati Chuprat: Conceptualization, Methodology, Writing – Original Draft Preparation, Project Administration. Evan Hanif Dwi Novianto: Visualization, Writing – Review & Editing. Akbariah Mohd Mahdzir: Data Curation, Writing – Review & Editing. Aizul Nahar Harun: Review & Editing. Yosiyuki Matsuura: Supervision, Funding Acquisition, Writing – Review & Editing.

Funding This research was partially supported by research grant funded by Yamaguchi University and Malaysia-Japan International Institute of Technology, Universiti Teknologi Malaysia (Grant Number: 4B823).

Data availability The analysis result presented in Figs. 2, 3, 4, 5, 6, and 7 and Table 3 was based on the extracted data from 169 selected papers as mentioned in the methodology section. The file that stored the extracted data is not attached in this paper, however, will be shared based on request.

Declaration

Conflicts of interest All authors declared that they have no conflicts of interest.

References

- Aaldering LJ, Leker J, Song CH (2019) Competition or collaboration? – Analysis of technological knowledge ecosystem within the field of alternative powertrain systems: a patent-based approach. *J Clean Prod* 212:362–371. <https://doi.org/10.1016/j.jclepro.2018.12.047>
- Abbas, A., Zhang, L., & Khan, S. U. (2014). A literature review on the state-of-the-art in patent analysis. In *World patent information* (Vol. 37, pp. 3–13). Elsevier Ltd. <https://doi.org/10.1016/j.wpi.2013.12.006>
- Abood A, Feltenberger D (2018) Automated patent landscaping. *Artif Intell Law* 26(2):103–125. <https://doi.org/10.1007/s10506-018-9222-4>
- Adams C, Tate D (2009) Computer-aided TRIZ ideality and level of invention estimation using natural language processing and machine learning. In: Tan R, Cao G, León N (eds) *Growth and development of computer-aided innovation*. Springer, Berlin, pp 27–37
- Adel ME, Harrison C (2024) Unravelling technology meta-landscapes: a patent analytics approach to assess trajectories and fragmentation. *World Patent Inf* 76:102256. <https://doi.org/10.1016/j.wpi.2023.102256>
- Ahn SJ, Yoon HY, Lee YJ (2021) Text mining as a tool for real-time technology assessment: application to the cross-national comparative study on artificial organ technology. *Technol Soc*. <https://doi.org/10.1016/j.techsoc.2021.101659>
- Alderucci D, Sicker D (2019) Applying artificial intelligence to the patent system. *Technol Innov* 20(4):415–425
- Altuntas S, Aba S (2022) Technology forecasting of unmanned aerial vehicle technologies through hierarchical S-curves. *Defence Sci J* 72(1):18–29
- Altuntas S, Erdogan Z, Dereli T (2020) A clustering-based approach for the evaluation of candidate emerging technologies. *Scientometrics* 124(2):1157–1177. <https://doi.org/10.1007/s11192-020-03535-0>
- Alvarez-Meaza I, Zarrabeitia-Bilbao E, Rio-Belver RM, Garechana-Anacabe G (2020) Fuel-cell electric vehicles: Plotting a scientific and technological knowledge map. *Sustainability* (Switzerland). <https://doi.org/10.3390/su12062334>
- Ameye N, Bughin J, van Zeebroeck N (2023) How uncertainty shapes herding in the corporate use of artificial intelligence technology. *Technovation* 127:102846. <https://doi.org/10.1016/j.technovation.2023.102846>

- Ampornphan P, Tongngam S (2020) Exploring technology influencers from patent data using association rule mining and social network analysis. *Information (Switzerland)*. <https://doi.org/10.3390/info11060333>
- Antunes L, Khorsandian F, Krabbe E, Sampson S, Wetherbee I (2018) Big data & public databases for patent research and analysis. <https://ipo.org/index.php/big-data-public-databases-for-patent-research-and-analysis/>
- Aristodemou L, Tietze F (2018) The state-of-the-art on intellectual property analytics (IPA): a literature review on artificial intelligence, machine learning and deep learning methods for analysing intellectual property (IP) data. *World Patent Inf* 55:37–51. <https://doi.org/10.1016/j.wpi.2018.07.002>
- Arnaldo Valdés RM, Burmaoglu S, Tucci V, da Costa B, Campos LM, Mattera L, Gomez Comendador VF (2019) Flight path 2050 and ACARE goals for maintaining and extending industrial leadership in aviation: a map of the aviation technology space. *Sustainability* 11(7):2065. <https://doi.org/10.3390/su11072065>
- Bastian M, Heymann S, Jacomy M (n.d.). Gephi: an open source software for exploring and manipulating networks visualization and exploration of large graphs. www.aaaai.org
- Birkle C, Pendlebury DA, Schnell J, Adams J (2020) Web of science as a data source for research on scientific and scholarly activity. *Quant Sci Stud* 1(1):363–376. https://doi.org/10.1162/qss_a_00018
- Bonino D, Ciaramella A, Corno F (2010) Review of the state-of-the-art in patent information and forthcoming evolutions in intelligent patent informatics. *World Patent Inf* 32(1):30–38. <https://doi.org/10.1016/j.wpi.2009.05.008>
- Bonner S, Kureshi I, Brennan J, Theodoropoulos G (2017) Exploring the evolution of big data technologies. In: Mistrik I, Bahsoon R, Ali N, Heisel M, Maxim B (eds), *Software architecture for big data and the cloud*. <https://doi.org/10.1016/B978-0-12-805467-3.00014-4>
- Brynjolfsson E, McAfee A (2017) Artificial intelligence, for real. *Harvard Bus Rev* 1:1–31
- Candelin-Palmqvist H, Sandberg B, Mylly UM (2012) Intellectual property rights in innovation management research: a review. *Technovation* 32(9):502–512. <https://doi.org/10.1016/j.technovation.2012.01.005>
- Chao MH, Trappey AJC, Wu CT (2021) Emerging technologies of natural language-enabled chatbots: a review and trend forecast using intelligent ontology extraction and patent analytics. In: *Complexity*, vol 2021. Hindawi Limited. <https://doi.org/10.1155/2021/5511866>
- Chapman P, Clinton J, Kerber R, Khabaza T, Reinartz T, Shearer C, Wirth R (2000) CRISP-DM 10: Step-by-step data mining guide. *SPSS Inc*, 9(13), 1–73
- Chen L, Xu S, Zhu L, Zhang J, Lei X, Yang G (2020) A deep learning based method for extracting semantic information from patent documents. *Scientometrics* 125(1):289–312. <https://doi.org/10.1007/s11192-020-03634-y>
- Chiarello F, Cimino A, Fantoni G, Dell’Orletta F (2018) Automatic users extraction from patents. *World Patent Inf* 54:28–38. <https://doi.org/10.1016/j.wpi.2018.07.006>
- Chiu Y-J (2018) Mining patent big data to forecast enterprise performance. In: Nah FF-H, Xiao BS (eds) *HCI in business, government, and organizations*. Springer International Publishing, pp 687–698
- Cho Y, Han YJ, Hwang J, Yu J, Kim S, Lee C, Lee S, Yi KP (2021) Identifying technology opportunities for electric motors of railway vehicles with patent analysis. *Sustainability (switzerland)* 13(5):1–14. <https://doi.org/10.3390/su13052424>
- Choi D, Song B (2018) Exploring technological trends in logistics: topic modeling-based patent analysis. *Sustainability (Switzerland)*. <https://doi.org/10.3390/su10082810>
- Chung J, Ko N, Yoon J (2021) Inventor group identification approach for selecting university-industry collaboration partners. *Technol Forecast Soc Change*. <https://doi.org/10.1016/j.techfore.2021.120988>
- Comai A (2018) Beyond patent analytics: insights from a scientific and technological data mashup based on a case example. *World Patent Inf* 55:61–77. <https://doi.org/10.1016/j.wpi.2018.10.002>
- de Paulo AF, Ribeiro EMS, Porto GS (2018) Mapping countries cooperation networks in photovoltaic technology development based on patent analysis. *Scientometrics* 117(2):667–686. <https://doi.org/10.1007/s11192-018-2892-6>
- Dehghani M, Mashatan A, Kennedy RW (2021) Innovation within networks – patent strategies for blockchain technology. *J Bus Ind Mark* 36(12):2113–2125. <https://doi.org/10.1108/JBIM-05-2019-0236>
- Du W, Wang Y, Xu W, Ma J (2021) A personalized recommendation system for high-quality patent trading by leveraging hybrid patent analysis. *Scientometrics* 126(12):9369–9391. <https://doi.org/10.1007/s11192-021-04180-x>
- Dunphy SM, Herbig PR, Howes ME (1996) *The innovation funnel*.

- Ernst CS, Olschewski I, Neumann NR, Harter C, Eckstein L (2015) Supplier strategies 2025 - winning technologies for the CO2 challenge. In: 24th Aachen colloquium automobile and engine technology 2015.
- Feng S, Magee CL (2020) Technological development of key domains in electric vehicles: improvement rates, technology trajectories and key assignees. *Appl Energy*. <https://doi.org/10.1016/j.apenergy.2019.114264>
- Feng L, Li Y, Liu Z, Wang J (2020) Idea generation and new direction for exploitation technologies of coal-seam gas through recombinative innovation and patent analysis. *Int J Environ Res Public Health*. <https://doi.org/10.3390/ijerph17082928>
- Feng J, Liu Z, Feng L (2021) Identifying opportunities for sustainable business models in manufacturing: Application of patent analysis and generative topographic mapping. *Sustain Prod Consump* 27:509–522. <https://doi.org/10.1016/j.spc.2021.01.021>
- Fernandez V (2022) Innovative intensity in the mining industry: Evidence from patent families. *Resour Policy* 78:102805. <https://doi.org/10.1016/j.resourpol.2022.102805>
- Füller J, Hutter K, Wahl J, Bilgram V, Tekic Z (2022) How AI revolutionizes innovation management – perceptions and implementation preferences of AI-based innovators. *Technol Forecast Soc Change* 178:121598. <https://doi.org/10.1016/j.techfore.2022.121598>
- Garzaniti N, Tekic Z, Kukolj D, Golkar A (2021) Review of technology trends in new space missions using a patent analytics approach. In: *Progress in aerospace sciences*, Vol. 125. Elsevier Ltd. <https://doi.org/10.1016/j.paerosci.2021.100727>
- Geum Y, Kim M (2020) How to identify promising chances for technological innovation: keygraph-based patent analysis. *Adv Eng Inform*. <https://doi.org/10.1016/j.aei.2020.101155>
- Grant E, Van den Hof M, Gold ER (2014) Patent landscape analysis: a methodology in need of harmonized standards of disclosure. *World Patent Inf* 39:3–10. <https://doi.org/10.1016/j.wpi.2014.09.005>
- Guderian CC (2019) Identifying emerging technologies with smart patent indicators: the example of smart houses. *Int J Innov Technol Manag*. <https://doi.org/10.1142/S0219877019500408>
- Haefner N, Wincent J, Parida V, Gassmann O (2021) Artificial intelligence and innovation management: a review, framework, and research agenda. *Technol Forecast Soc Change*. <https://doi.org/10.1016/j.techfore.2020.120392>
- Hällgren M (2012) The construction of research questions in project management. *Int J Project Manage* 30(7):804–816. <https://doi.org/10.1016/J.IJPROMAN.2012.01.005>
- IBM (2021) CRISP-DM help overview. <https://www.ibm.com/docs/en/spss-modeler/saas?topic=dm-crisp-help-overview>
- Jackson TW, Farzaneh P (2012) Theory-based model of factors affecting information overload. *Int J f Inf Manag* 32(6):523–532. <https://doi.org/10.1016/j.ijinfomgt.2012.04.006>
- Jain R, Tripathi M, Agarwal V, Murthy J (2020) Patent data analytics for technology benchmarking: R-based implementation. *World Patent Inf*. <https://doi.org/10.1016/j.wpi.2020.101952>
- Jang H, Yoon B (2021) TechWordNet: Development of semantic relation for technology information analysis using F-term and natural language processing. *Inf Process Manag*. <https://doi.org/10.1016/j.ipm.2021.102752>
- Jang H, Jeong Y, Yoon B (2021) TechWord: development of a technology lexical database for structuring textual technology information based on natural language processing. *Expert Syst Appl*. <https://doi.org/10.1016/j.eswa.2020.114042>
- Jun S (2021) Technology integration and analysis using boosting and ensemble. *J Open Innov Technol Mark Complex* 7(1):27. <https://doi.org/10.3390/joitmc7010027>
- Jun S, Lee SJ, Ryu JB, Park S (2015) A novel method of IP R&D using patent analysis and expert survey. www.wipo.org
- Kang J, Lee J, Jang D, Park S (2019) A methodology of partner selection for sustainable industry-university cooperation based on LDA topic model. *Sustainability* 11(12):3478. <https://doi.org/10.3390/su11123478>
- Kang JN, Wei YM, Liu L, cui, & Wang, J. W. (2021) Observing technology reserves of carbon capture and storage via patent data: paving the way for carbon neutral. *Technol Forecast Soc Change*. <https://doi.org/10.1016/j.techfore.2021.120933>
- Khan MAD, Uddin MF, Gupta N (2014) Seven V's of big data understanding big data to extract value. In: *Proceedings of the 2014 Zone 1 conference of the American society for engineering education*, 1–5. <https://doi.org/10.1109/ASEEZone1.2014.6820689>
- Kim G, Bae J (2017) A novel approach to forecast promising technology through patent analysis. *Technol Forecast Soc Change* 117:228–237. <https://doi.org/10.1016/j.techfore.2016.11.023>

- Kim J, Lee S (2015) Patent databases for innovation studies: a comparative analysis of USPTO, EPO, JPO and KIPO. *Technol Forecast Soc Chang* 92:332–345. <https://doi.org/10.1016/j.techfore.2015.01.009>
- Kim GJ, Park SS, Jang DS (2015) Technology forecasting using topic-based patent analysis. *J Sci Ind Res* 74(5):265–270
- Kim KH, Han YJ, Lee S, Cho SW, Lee C (2019) Text mining for patent analysis to forecast emerging technologies in wireless power transfer. *Sustainability (Switzerland)*. <https://doi.org/10.3390/su11226240>
- Kim K, Park K, Lee S (2019b) Investigating technology opportunities: the use of SAOx analysis. *Scientometrics* 118(1):45–70. <https://doi.org/10.1007/s11192-018-2962-9>
- Kim J, Jeong B, Kim D (2021) Introduction. In: Kim J, Jeong B, D. Kim (Eds.), *Patent analytics: transforming IP strategy into intelligence*, Springer, Singapore. pp 1–8. https://doi.org/10.1007/978-981-16-2930-3_1
- Kitchenham B, Pearl Brereton O, Budgen D, Turner M, Bailey J, Linkman S (2009) Systematic literature reviews in software engineering – a systematic literature review. *Inf Softw Technol* 51(1):7–15. <https://doi.org/10.1016/j.infsof.2008.09.009>
- Kraus S, Breier M, Dasí-Rodríguez S (2020) The art of crafting a systematic literature review in entrepreneurship research. *Int Entrepreneurship Manag J* 16(3):1023–1042. <https://doi.org/10.1007/s11365-020-00635-4>
- Krestel R, Chikkamath R, Hewel C, Risch J (2021) A survey on deep learning for patent analysis. *World Patent Inf*. <https://doi.org/10.1016/j.wpi.2021.102035>
- Lee M, He G (2021) An empirical analysis of applications of artificial intelligence algorithms in wind power technology innovation during 1980–2017. *J Clean Prod*. <https://doi.org/10.1016/j.jclepro.2021.126536>
- Lee C, Kim J, Kwon O, Woo H-G (2016a) Stochastic technology life cycle analysis using multiple patent indicators. *Technol Forecast Soc Chang* 106:53–64. <https://doi.org/10.1016/j.techfore.2016.01.024>
- Lee J, Kim G, Park S, Jang D (2016) Hybrid corporate performance prediction model considering technical capability. *Sustainability (Switzerland)*. <https://doi.org/10.3390/su8070640>
- Lee C, Kwon O, Kim M, Kwon D (2018) Early identification of emerging technologies: a machine learning approach using multiple patent indicators. *Technol Forecast Soc Chang* 127:291–303. <https://doi.org/10.1016/j.techfore.2017.10.002>
- Lee JW, Lee WK, Sohn SY (2021) Patenting trends in biometric technology of the Big Five patent offices. *World Patent Inf*. <https://doi.org/10.1016/j.wpi.2021.102040>
- Lee S, Hwang J, Cho E (2022) Comparing technology convergence of artificial intelligence on the industrial sectors: two-way approaches on network analysis and clustering analysis. *Scientometrics* 127(1):407–452. <https://doi.org/10.1007/s11192-021-04170-z>
- Liu Y, Mai S (2024) The evaluation of innovation ability of China's information technology application innovation enterprises based on cloud modeling. *Finance Res Lett* 60:104894. <https://doi.org/10.1016/j.frl.2023.104894>
- Liu Y, Xu X (2017) Industry 4.0 and cloud manufacturing: a comparative analysis. *J Manuf Sci Eng Trans ASME* 10(1115/1):4034667
- Liu Y, Yan Z, Cheng Y, Ye X (2018) Exploring the technological collaboration characteristics of the global integrated circuit manufacturing industry. *Sustainability (Switzerland)*. <https://doi.org/10.3390/su10010196>
- Liu B, Lai M, Wu JL, Fu C, Binaykia A (2020) Patent analysis and classification prediction of biomedicine industry: SOM-KPCA-SVM model. *Multimedia Tools Appl* 79(15–16):10177–10197. <https://doi.org/10.1007/s11042-019-7422-x>
- Liu Y, Wang G, Zhou Y, Liu Y (2022) Advanced technology evolution pathways of nanogenerators: a novel framework based on multi-source data and knowledge graph. *Nanomaterials*. <https://doi.org/10.3390/nano12050838>
- Mao G, Han Y, Liu X, Crittenden J, Huang N, Ahmad UM (2022) Technology status and trends of industrial wastewater treatment: a patent analysis. *Chemosphere*. <https://doi.org/10.1016/j.chemosphere.2021.132483>
- Maragakis M, Rouni MA, Mouza E, Kanetidis M, Argyrakakis P (2023) Tracing technological shifts: time-series analysis of correlations between patent classes. *Eur Phys J plus* 138(9):776. <https://doi.org/10.1140/epjp/s13360-023-04344-z>

- Miao H, Wang Y, Li X, Wu F (2022) Integrating technology-relationship-technology semantic analysis and technology roadmapping method: a case of elderly smart wear technology. *IEEE Trans Eng Manage* 69(1):262–278. <https://doi.org/10.1109/TEM.2020.2970972>
- Minghan S, Xiuzhu Z, Man J (2020) Exploring the innovation landscape of bamboo fiber technologies from global patent data perspective. *Cellulose* 27(16):9137–9156
- Moehrle MG, Walter L, Bergmann I, Bobe S, Skrzypale S (2010) Patinformatics as a business process: a guideline through patent research tasks and tools. *World Patent Inf* 32(4):291–299. <https://doi.org/10.1016/j.wpi.2009.11.003>
- Moher D, Liberati A, Tetzlaff J, Altman DG (2009) Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *BMJ (Online)* 339(7716):332–336. <https://doi.org/10.1136/bmj.b2535>
- Naeini AB, Zamani M, Daim TU, Sharma M, Yalcin H (2022) Conceptual structure and perspectives on “innovation management”: a bibliometric review. *Technol Forecast Soc Change*. <https://doi.org/10.1016/j.techfore.2022.122052>
- Nordensvard J, Zhou Y, Zhang X (2018) Innovation core, innovation semi-periphery and technology transfer: the case of wind energy patents. *Energy Policy* 120:213–227. <https://doi.org/10.1016/j.enpol.2018.04.048>
- Panetta K (2018) Gartner top 10 strategic technology trends for 2018. <https://www.gartner.com/smarterwithgartner/gartner-top-10-strategic-technology-trends-for-2018>
- Park Y-N, Lee Y-S, Kim J-J, Lee TS (2018) The structure and knowledge flow of building information modeling based on patent citation network analysis. *Autom Construct* 87:215–224. <https://doi.org/10.1016/j.autcon.2017.12.023>
- Park H, Anderson TR, Seo W (2021) Regional innovation capability from a technology-oriented perspective: an analysis at industry level. *Comput Ind*. <https://doi.org/10.1016/j.compind.2021.103441>
- Puccetti G, Chiarello F, Fantoni G (2021) A simple and fast method for named entity context extraction from patents. *Expert Syst Appl*. <https://doi.org/10.1016/j.eswa.2021.115570>
- Raguseo E (2018) Big data technologies: An empirical investigation on their adoption, benefits and risks for companies. *Int J Inf Manag* 38(1):187–195. <https://doi.org/10.1016/j.ijinfomgt.2017.07.008>
- Sasaki H, Sakata I (2021) Identifying potential technological spin-offs using hierarchical information in international patent classification. *Technovation*. <https://doi.org/10.1016/j.technovation.2020.102192>
- Schröer C, Kruse F, Gómez JM (2021) A Systematic literature review on applying CRISP-DM process model. *Procedia Comput Sci* 181:526–534. <https://doi.org/10.1016/j.procs.2021.01.199>
- Shen Y-C, Wang M-Y, Yang Y-C (2020) Discovering the potential opportunities of scientific advancement and technological innovation: a case study of smart health monitoring technology. *Technol Forecast Soc Change* 160:120225. <https://doi.org/10.1016/j.techfore.2020.120225>
- Shi X, Zhang Q (2018) Inbound open innovation and radical innovation capability: the moderating role of organizational inertia. *J Organ Chang Manag* 31(3):581–597. <https://doi.org/10.1108/JOCM-07-2017-0262>
- Shibata N, Kajikawa Y, Takeda Y, Matsushima K (2008) Detecting emerging research fronts based on topological measures in citation networks of scientific publications. *Technovation* 28(11):758–775. <https://doi.org/10.1016/j.technovation.2008.03.009>
- Smojver V, Štorga M, Potočki E (2019) Determining the life cycle phase of a technology based on patent data. *Tehnicki Vjesnik* 26(1):222–229
- Snyder H (2019) Literature review as a research methodology: an overview and guidelines. *J Bus Res* 104:333–339. <https://doi.org/10.1016/j.jbusres.2019.07.039>
- Song B, Suh Y (2019) Identifying convergence fields and technologies for industrial safety: LDA-based network analysis. *Technol Forecast Soc Chang* 138:115–126. <https://doi.org/10.1016/j.techfore.2018.08.013>
- Tekic Z, Füller J (2023) Managing innovation in the era of AI. *Technol Soc* 73:102254. <https://doi.org/10.1016/j.techsoc.2023.102254>
- Thavorn J, Muangsin V, Gowanit C, Muangsin N (2021) A scientometric assessment of agri-food technology for research activity and productivity. *Publications*. <https://doi.org/10.3390/publications9040057>

- Trappey AJC, Trappey C, v., Wu, J. L., & Wang, J. W. C. (2020) Intelligent compilation of patent summaries using machine learning and natural language processing techniques. *Adv Eng Inform.* <https://doi.org/10.1016/j.aei.2019.101027>
- Trippe A (2015) Guidelines for preparing patent landscape reports. WIPO.
- Walter L, Denter NM, Keibel J (2022) A review on digitalization trends in patent information databases and interrogation tools. *World Patent Inf.* <https://doi.org/10.1016/j.wpi.2022.102107>
- Walter L, Denter NM, Keibel J (2022b) A review on digitalization trends in patent information databases and interrogation tools. In: *World patent information*, vol 69. Elsevier Ltd. <https://doi.org/10.1016/j.wpi.2022.102107>
- Wang C, Wood LC, Abdul-Rahman H, Lee YT (2016) When traditional information technology project managers encounter the cloud: Opportunities and dilemmas in the transition to cloud services. *Int J Project Manag* 34(3):371–388. <https://doi.org/10.1016/j.ijproman.2015.11.006>
- Xu X, Gui M (n.d.) Applying data mining techniques for technology prediction in new energy vehicle: a case study in China. <https://doi.org/10.1007/s11356-021-15298-z/Published>
- Yalcin H, Daim T (2021) Mining research and invention activity for innovation trends: case of blockchain technology. *Scientometrics* 126(5):3775–3806. <https://doi.org/10.1007/s11192-021-03876-4>
- Yanhui S, Lixin L (2024) Inventor bibliographic-patent-coupling analysis and inventor-patent-classification-coupling analysis: a comparative analysis based on NPE. *Scientometrics* 129(2):745–765. <https://doi.org/10.1007/s11192-023-04713-6>
- Yoshikuni AC, Dwivedi R, Zhou D, Wamba SF (2023) Big data and business analytics enabled innovation and dynamic capabilities in organizations: Developing and validating scale. *Int J Inf Manag Data Insights* 3(2):100206. <https://doi.org/10.1016/j.ijime.2023.100206>
- Yu J, Hwang JG, Hwang J, Jun SC, Kang S, Lee C, Kim H (2020) Identification of vacant and emerging technologies in smart mobility through the gtm-based patent map development. *Sustainability (Switzerland)* 12(22):1–22. <https://doi.org/10.3390/su12229310>
- Yuan X, Li X (2021) Mapping the technology diffusion of battery electric vehicle based on patent analysis: a perspective of global innovation systems. *Energy.* <https://doi.org/10.1016/j.energy.2021.119897>
- Zhang J, Yan Y, Guan J (2019) Recombinant distance, network governance and recombinant innovation. *Technol Forecast Soc Change* 143:260–272. <https://doi.org/10.1016/j.techfore.2019.01.022>
- Zhang HE, Wong KH, Chang V (2021) Patent analysis in the 5G network. *J Glob Inf Manag* 29(6):1–28. <https://doi.org/10.4018/jgim.20211101.0a28>
- Zhao Q (2018) Green innovation regarding electromobility in Germany and China: who and how to get involved? *J Scientometric Res* 7(1):1–8. <https://doi.org/10.5530/jscires.7.1.1>
- Zhou Y, Pan M, Urban F (2018) Comparing the international knowledge flow of China's wind and solar photovoltaic (PV) industries: Patent analysis and implications for sustainable development. *Sustainability (Switzerland)*. <https://doi.org/10.3390/su10061883>
- Zhou Y, Li Z, Liu Y, Deng F (2021) Network proximity and communities in innovation clusters across knowledge, business, and geography: evidence from china. *IEEE Trans Eng Manage* 68(5):1388–1397. <https://doi.org/10.1109/TEM.2020.3032160>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.