ABSTRACT

Innovation is a cornerstone of progress, and patents play a pivotal role in protecting and promoting novel ideas and technologies. This paper delves into the dynamic intersection of text mining and innovation studies by focusing on patent data. This literature review paper presents an in-depth assessment of the trends and developments in text mining techniques as they relate to patent analysis within the context of innovation studies. This paper presents a thorough study of text mining literature through a systematic literature review of 162 articles in 20 peer-reviewed journals published from 2003 to 2022. We undertake a thorough literature review in this study, identifies themes and categorizes the papers into five clusters of Technology opportunities identification, Firm competition strategies, Emerging technology forecasting and evaluation, Patent technical intelligence, Technological convergence and open-endedness. The study shows that most articles concentrated on technological firms and focused on technological-driven strategies. This paper highlights key results, open issues and outlines a compelling research agenda for further investigation and development in the field of text mining in innovation studies related to patents.

INTRODUCTION

Innovation studies is a relatively new and rapidly expanding discipline of social sciences mostly influenced by the works of Schumpeter (Nelson and Winter, 1983; Fagerberg et al, 2009). The field of innovation studies is continually evolving, driven by the dynamic landscape of technological advancements and their profound impact on industries and societies. The application of text mining analysis to patent-related research has become a crucial field of study in this dynamic field (Peng, 2018). This multidisciplinary approach provides a window into the complexities of innovation processes, illuminating trends, obstacles, and a research agenda for the future that has the potential to fundamentally alter our comprehension of innovation in the digital era. In this paper, we explore the field of Text Mining Analysis of Patents in Innovation Studies, aiming to identify the dominant patterns, identify relevant issues, and determine the direction of future studies in this fascinating area. Knowledge discovery from text (KDT), which includes text mining, has several applications. More and more software packages for tasks as varied as risk management, corporate analytics, customer service, fraud detection, and social media use this adaptable method. It has wide-ranging uses in fields as diverse as medicine, business, education, and even social media. In essence, text mining involves the extraction of valuable insights and information from textual sources, encompassing structured data, semi-structured data (e.g., XML and JSON), and unstructured text resources, as detailed by Kumari et al. (2021). The origins of text mining trace back to the work of Feldman et al., (1998). Within the domain of patent analysis, several recent surveys shed light on various facets. Krestel et al. (2021) studied deep learning techniques in text mining analysis. They gave an overview of datasets, text representation methods, and deep neural network architectures used in different patent analysis tasks. Meanwhile, Ozcan and Islam (2017) embarked on a descriptive journey through patent literature, focusing on the search requirements essential for information retrieval, systems, and applications. Their study aimed to discern the overarching needs of patent users concerning search functional requirements. However, we present a systematic review on the existing body of knowledge in text mining, Nambisan et al. (2017) made a notable observation: the widespread adoption of digital technologies has not only transformed the essence of innovation but has also revolutionized the way we analyze innovation processes and their outcomes. Numerous academics from various business-related subfields have agreed with this recognition of the transformative power of digital tools in innovation research (George et al., 2014; Chintagunta et al., 2016; Antons and Breidbach, 2018). Li et al. (2019) contributes significantly with a theoretical framework that embraces the fusion of science and technology, leveraging text mining and expert evaluation. Their work harnesses data from patents and scientific articles to forecast technological trends. In a more recent study, Changyong Lee (2021) explores the field of data analytics in technological forecasting, examining publications in esteemed journals within the technology and innovation management domain. The study introduces a process-focused morphological matrix, which provides a lucid yet comprehensive perspective, enabling a thorough exploration of the full spectrum of data analytics applied to technological forecasting.
The majority of scientific papers concentrate on particular text mining techniques for information extraction from text documents in innovation studies and emphasize the use of different text mining algorithms on unstructured data. However, a comprehensive examination of the various text mining techniques and cluster analysis is still absent. It is against these backdrops that, this study aims to offer a thorough review of the literature on text mining applications in innovation studies. We identified and reviewed a set of 162 articles on innovation-related studies published in a collection of 20 prestigious innovation-related journals for the past two decades. The study surveys and analyzes numerous studies and practices, providing readers with a comprehensive understanding of how text mining techniques are evolving and being applied to innovation research. Second, the paper provides a foundation for future research endeavors in the field of text mining in innovation studies. In addition, it provides a road map for academicians and researchers interested in advancing the field by outlining a structured research agenda that highlights prospective directions for future investigation.

In this context, we ask; what are the key innovation focus areas of the published papers? What are the main text mining methods employed by the examined papers? Which industries dominate the examined papers? Therefore, by suggesting a set of recommended practices, we provide practical expertise on how text mining is used in innovation studies. Therefore, the study aims to answer these questions through a systematic literature review. The study discovered that most text mining analyses in innovation studies uses case study analysis, making it difficult for researchers to extrapolate findings to other contexts. This paper discusses the literature review related to text mining in innovation research. The researchers identified and reviewed 162 articles that have been published in peer review in innovation and management journals from 2003 to 2022. We tabulated the papers into clusters, pinpointing their technological focus areas, main text mining methods and tools with their years of publication and conclusions. The paper finishes with key findings and suggestions for further research agenda. To the best of our knowledge and considering the growing interest in the field of innovation studies, no survey article has focused on this direction. This research, as far as we know, is the first stream to survey this direction. The rest of the paper is structured as follows. In the second section, the study outlines the literature review in section 2. The methodology employed in the review is presented in section 3. Sections 4 highlights the main results from the review of literature. The conclusions are presented in section 5. Section 6 contains the future research avenues and directions.

LITERATURE REVIEW
Text mining, a subset of natural language processing (NLP), has emerged as a powerful tool in the field of innovation studies. Its application extends beyond traditional data analysis methods, offering unique capabilities in extracting valuable insights from unstructured textual data, particularly in the context of patents and innovation-related documents. Text mining allows researchers to uncover hidden knowledge within vast volumes of textual data. In innovation studies, this means identifying emerging trends, technological advancements, and novel ideas by analyzing patents, research papers, and other innovation-related texts (Pantano & Stylidis 2021). Text mining techniques, such as information extraction and keyword analysis, streamline the process of information retrieval. Patents and innovation documents are rich sources of qualitative information. Text mining bridges the gap by quantifying qualitative data and converts textual content into structured data, making it amenable to statistical analysis (Schmiedel et al., 2019). Text mining techniques, including clustering and topic modeling, aid in grouping similar documents or concepts. A recent study by Choi et al. (2021) on emerging technologies and future ecosystem contends that in the face of swift technological advancements and dynamic shifts in business value systems, it has become important for organizations to pinpoint nascent and promising technologies capable of effectively addressing external disruptive factors. These technologies serve as the catalysts for launching new ventures or enhancing existing ones. Researchers can discern patterns of technological convergence and divergence, fostering a deeper understanding of innovation dynamics. Text mining bridges the gap between technology and social sciences, encourages interdisciplinary collaboration between computer scientists, linguists, and innovation scholars (Rhodes et al., 2022). Such collaboration fosters a holistic approach to innovation studies, incorporating both quantitative and qualitative dimensions. Text mining techniques for patent analysis are indispensable tools in the modern innovation landscape. With the use of software that can find concepts, patterns, subjects, keywords, and other properties in the data, text mining is the exploration and analysis of huge amounts of unstructured text data (Lydia et al., 2020). Text mining techniques have been used significantly more in recent years in a variety of research fields, including new product creation, security applications, sentiment analysis, online media applications, biomedical applications, business and marketing application, digital humanities and computational sociology, and so on. Text mining techniques, including information extraction, topic tracking, topic summaries, classification, clustering, association rule mining (ARM), sentiment analysis, etc., are employed to lessen the amount of manual labor required to analyze unstructured, long, and rich textual data. These techniques facilitate the extraction of valuable insights from the enormous and complex corpus of patent documents by leveraging the power of natural language processing and machine learning (Olivetti et al., 2022). These insights go beyond mere keyword searches, delving into the nuanced relationships between concepts.
Table 1: Journal statistics

<table>
<thead>
<tr>
<th>Name of Journal</th>
<th>Frequency</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scientometrics</td>
<td>52</td>
<td>32%</td>
</tr>
<tr>
<td>Technological Forecasting &amp; Social Change</td>
<td>46</td>
<td>28%</td>
</tr>
<tr>
<td>Technology analysis and strategic management</td>
<td>12</td>
<td>7%</td>
</tr>
<tr>
<td>Technovation</td>
<td>9</td>
<td>6%</td>
</tr>
<tr>
<td>Journal of Informatics</td>
<td>9</td>
<td>6%</td>
</tr>
<tr>
<td>IEEE Transactions on Engineering Management</td>
<td>8</td>
<td>5%</td>
</tr>
<tr>
<td>Research and Development Management</td>
<td>8</td>
<td>5%</td>
</tr>
<tr>
<td>Research Policy</td>
<td>5</td>
<td>3%</td>
</tr>
<tr>
<td>Economics of Innovation and New Technology</td>
<td>2</td>
<td>1%</td>
</tr>
</tbody>
</table>

Technologies, and inventors. Patent analysis using text mining facilitates the identification of emerging trends and technologies, providing a crucial competitive advantage to businesses and researchers. Furthermore, it streamlines the prior art search process, aiding inventors in avoiding patent infringements and ensuring the novelty of their inventions. According to Trappey et al. (2017 & 2015) text mining techniques enable the creation of comprehensive patent landscapes, helping organizations make informed decisions about research and development investments. Patent analysis is an indispensable pillar of innovation, offering a wealth of information and insights that fuel technological advancements, guide strategic decisions, and underpin effective intellectual property management. It enables proactive technology monitoring, helping identify emerging trends and competitive dynamics (Bharadiya 2023). Inventors benefit from thorough prior art searches, enhancing the efficiency of the patent system. Businesses and research institutions use patent analysis to formulate innovation strategies, guiding resource allocation and partnerships (Igartua et al., 2010). Analysts leverage it for technological forecasting, predicting future trends, and gaining a competitive edge. Policymakers utilize patent analysis to inform innovation policies, fostering economic growth and competitiveness. Additionally, it facilitates mapping innovation ecosystems, identifying key players, and enhancing collaboration.

**METHODOLOGY**

The study adopted a systematic literature review as outlined by Palmatier et al. (2018) and Kraus et al. (2022). The procedure included three steps: 1) Planning the review, 2) selection and extraction of data and 3) reporting of findings. The last step is presented in the results section of the study.

**Planning the Review**

In this study, we followed a specific sequence of steps. We started the planning process by gathering and organizing important articles from Web of Science database system. The decision to opt for the Association of Business Schools (ABS) list stems from the fact it has greater comprehensiveness compared to alternative journal ranking lists such as Social Sciences Citation Index (SSCI) and Scopus. Initially, we use the database’s advanced search features to narrow down our results based on publication year and specific journals related to innovation studies and text mining in patents. We conducted a search on keywords, titles, and abstracts of published articles to download the relevant research literature published from 2003 to 2022 in 20 peer reviewed journals. This initial search yielded 546 research articles. To perform metric analysis, we utilized the Web of Science (WOS) indexed Journal Ranking database, focusing on journal citations within the domains of economics and management. Employing keyword searches, we refined our exploration to encompass specific areas such as ‘text mining’ ‘patent analysis’ ‘patent text analysis’ ‘patent data mining’ ‘intellectual property analysis’. This targeted approach facilitated the identification of important articles and research within this specialized field.

**Selection and Extraction**

Again, we focused only on publications that were relevant to their study (all 1-star ranking journals in ABS spectrum were excluded). As a result, the total number of potential contributions was reduced further. We considered papers with empirical content focus on text mining and patent analysis hence concentrating exclusively on the titles, abstract and keywords of the remaining articles to exclude papers that were parallel to the scope and objectives of the study. This brought down the number of research papers to 162. The important details, including title, abstract, keywords, authors names and affiliations, journal name, year of publication, and number of citations, were extracted and exported into an MS Excel spreadsheet. Subsequently, a thorough assessment of the titles and abstracts was conducted to exclude articles that were not relevant to this study.

It is found that articles related to patent text analysis are mainly concentrated in two journals, Scientometrics and Technological Forecasting & Social Change, with 52 and 46 papers respectively. With regards to the rest of the journals, 12 papers in Technology analysis and strategic management, 9 each in journals of technovation and informatics. The rest of the information is shown in Table 1 below which summarizes the particular journals we searched.
RESULTS
Main Methods and Indicators of the Examined Papers
Within the realm of patent analysis, a number of text mining methodologies unfolds, each offering unique insights into the complex web of technological innovation. One such method gaining prominence is the Subject-Action-Object (SAO) structure-semantic mining approach. This method involves the extraction of SAO structures from patent abstracts using text mining techniques. These structures serve as a foundation for mining semantic information embedded in patent texts related to emerging technologies. SAO proves to be a valuable tool, illuminating key technical components and enabling academics to channel their creativity into other domains. A notable 13% of the reviewed studies employ SAO-based semantic techniques to identify fundamental technological components within areas of interest. Semantic similarity algorithms are employed to cluster patent texts, and SAO structure similarities are leveraged to trace the evolution of technology development and deployment trajectories.

In addition, there are a number of other methods including citation analysis employed by these papers. By measuring the number of times, a certain author, article, or publication has been cited by other works, a technique known as citation analysis can be used to assess their relative relevance or impact. Citation data offers citation relationships that may be used to study technology diffusion, value, or effect across several patents. Many studies in the firm competitive cluster have utilized patent citation analysis to build information exchange networks for quantifying data moves. For instance, No et al. (2015) defined technology-based Business Model patents as knowledge flow drivers and quantified the degree of knowledge flow generated by technology-based Business Models using patent citation and text data.

Link prediction describes the evolution of the node associations as well as the influences on node associations. It is a technique for predicting the possibility of a future connection between two nodes in a network. Many scientific disciplines such as Medicine (Yoon et al., 2018), 3D printing technology (Han et al., 2021) and Water purification methods (Yoon et al., 2018) have used link prediction analysis. There are several possible uses for link prediction in social networks, including suggesting new products to users, meeting people, and spotting fictitious relationships. According to these studies, Link prediction can offer insight into future technological convergence, aiding prompt decision-making in developing technologies, if it is incorporated into objective and credible dataset. Another text mining method that was featured in this review is Network analysis. Network analysis enables us in fully comprehending the social network dynamic relationship as well as the structure or process of change in natural phenomena. Identifying the most important node in a network is key in network analysis. According to Sun et al. (2023), network Analysis can assist researchers visualize the network link and communicate the results of the investigation. Most significantly. Network Analysis may uncover hidden patterns those standard qualitative measurements may miss, as well as aid experts in identifying upcoming development trends of new technologies. Network analysis is beneficial for the quantitative and visual interpretation of human association analysis.

Morphological Analysis (MA) is a technique for locating, organizing, and researching the whole collection of potential connections present in a particular multidimensional issue complex. Morphological Analysis has been effectively used in several fields, including control of technical development and modeling the bioethics of drug redevelopment, in strategic planning and decision assistance. As used in their study, Yoon et al., (2008) stated that Morphological Analysis is typically used to organize an issue by breaking it down into subsystems, identifying the morphology of existing products and technology, and so providing innovative opportunities for roadmap development. Breaking down morphologically complicated words into their distinct morphemes is known as morphological analysis, also referred as structural analysis. Morphological analysis is the initial stage of text preprocessing. Co-word analysis, which many researchers have employed in their research, is a content analysis technique that combines bibliometrics with text mining technology to discover the hidden meaning of texts. Some social researchers utilize co-word analysis to

| Journal of Economics & Management Strategy | 2 | 1% |
| Strategic Management Journal | 2 | 1% |
| Industrial and Corporate Change | 1 | 1% |
| International Journal of Technology Management | 1 | 1% |
| Journal of Innovation and Knowledge | 1 | 1% |
| Technology in Society | 1 | 1% |
| Journal of Engineering and Technology Management | 1 | 1% |
| Journal of Knowledge Management | 1 | 1% |
| Journal of the Association for Information Science and Technology | 1 | 1% |
| University of Chicago Journal | 1 | 1% |
| R&D Management | 1 | 1% |
| Total | 162 | 100 |

https://journals.e-palli.com/home/index.php/ajebi
examine the growth and organization of the academic literature on gender inequalities in science and higher education. Lee et al., (2016) employed co-word analysis in their study of using patent information for designing new product and technology. This section describes the various text mining techniques used in patent analysis in the papers under review. The selected articles have been reviewed and analyzed based on the main text mining methods employed. Figure 2 shows the various main text mining methods of the selected papers in this systematic review with their corresponding scores as displayed.

**Industry of the Examined Papers**

Figure 2 illustrates the distribution of scores of the studies and their various industries. As indicated below, technological firms accounted for the highest score of 16%, patent and photovoltaics scored 10% each, business strategy and biochemical scored 8% each of the reviewed papers. Also, Pharmaceuticals ICT and electronic industry accounted for 7% each, automobile sector scored 5%. The rest of the industries, AI, medical industry, legal system, wireless power firms, Aerospace industry, construction, Fuel & Solar cells, SMEs all scored less than 5%.

**Clusters**

We clustered the articles into five thematic areas of; technology opportunities identification (48), firm competition strategies (22), emerging technology forecasting and evaluation (27), patent technical intelligence (37), technological convergence and open-
Cluster 1: Technology Opportunity Identification
Identification of technological opportunities is the process of identifying potential ways to use technology to improve the production or use of products (Cho et al., 2013). The majority of publications in this cluster are focused on discovery of new opportunities based on patents (Song et al., 2017; Jang et al., 2021). Some of the reviewed articles (Lee et al., 2020; Yoon and Park, 2005) propose a methodology for determining if potential emerging technologies will expand rapidly and have a significant influence on social and technological domains in the future. Thus, their publications included a sampling of “expert views regarding the future,” i.e., remarks from professionals focused on the near future from both general and specialized technological groups. To identify untapped technological areas and to outline the precise course of technological advancement, Teng et al. (2021) proposed a four-stage approach to patent text data in their study of the discovery of proton exchange membrane fuel cells based on generative topographic mapping. Three cutting-edge food processing innovations include fuel cells based on generative topographic mapping. Subject-Action-Object (SAO) was employed by other studies to identify core technological components and has been used as a useful tool in technological mining. The subject-action-object (SAO)-based semantic patent analysis was suggested in many of the reviewed papers as a technique for identifying new technological opportunities (Yoon & Kim, 2011; Wang et al., 2017; Yang et al., 2017; Yoon and Kim, 2012; Choi et al., 2011). A study by the former, Yoon et al. (2011) used outlier detection to find outlier patents in a particular technology field that were unusual. The study concentrated his studies and analysis on identifying technological competition trends using different fields as case studies. Yun et al. (2021) investigated the value of expired patents and argue that the distinctive qualities of lapsed patents as opportunities have been generally overlooked. Their proposed method is applied to bio cosmetics products. A section of the studies focused on technology opportunity analysis, although the majority of them have been narrowly focused on the discovery of new technological concepts. According to the technological capabilities built into their current product, Lee et al. (2020) used a product landscape analysis to identify product areas across various disciplines into which businesses might expand. In general, patent information can give people a wealth of technical and business information to aid in the development of new concepts and the planning of specialized technological fields. For instance, Liu and Luo’s (2008) paper on the gait of a biped humanoid robot was used as an illustration to look into the relative research capacities and patent citation needs for patent owners and patent mappers. Based on the examination of these cluster of papers, it was observed that some scholars such Yoon and Kim (2011) used semantic analysis to identify textual commonalities that allow for pairwise document comparison. This technique was employed by the studies because it enables papers to be represented as a combination of ideas, making it easier to detect documents that are similar or distinct. Further analyses of this cluster of papers lead us to conclude that most of the publications provides further information and analyses to pinpoint areas where there are patent gaps and technological hot spots. In summary, this cluster underscores the significance of patent analysis and text mining techniques in the identification of technological opportunities, providing valuable insights for innovation and business strategy. It emphasizes the need to consider both valid and expired patents, utilize SAO-based semantic analysis, and leverage patent information for informed decision-making in technology-driven industries.

Cluster 2: Firm Competition Strategies
The articles in this cluster place particular attention on issues related to patent roadmap for firm competition analysis and strategy planning. A firm competitive strategy is a long-term action plan developed by a firm to gain a competitive edge over its competitors in the industry. In their study, Wang et al. (2014) extend a traditional Latent Dirichlet Allocation (LDA) for patent competitive intelligence analysis. The latent associations of the collected technology words are used to uncover underlying topic structures using the extended LDA model. Merger and Acquisition (M&A) also appeared in this cluster as a strategy for enhancing technological capabilities of firms. Park et al. (2013) put forth a framework to help M&A target selection decision-makers identify and assess companies from a technological perspective. Comparable to this, Qi et al. (2022) developed a methodical framework based on topic analysis and link prediction that investigates the process of selecting collaborators for cooperative creativity. This cluster also explores knowledge flows to have a better understanding of technological driven based business models. Some of the studies addresses the managerial aspects of business method patenting (Moehrle et al., 2018; Lee et al., 2013; No et al., 2015). Moehrle et al. (2018) divided the 37,000 RFID-related patents from the 1990s to 2014 into technological and business method patents using a case study of Radio Identification Devices (RFID). Using morphological analysis as the foundation of their suggested methodology, Lee et al. (2013) structured various business model kinds. Their research suggested a dynamic patent analysis that might reveal intricate connections between business method patents and
show patterns in the development of technology-driven business models. Their study, however, was limited to only one business area, electronic shopping, so the case study conclusions cannot be extended to other businesses.

In order to understand how the key skills of one organization are over time mirrored in the innovation activities of another, Kronemeyer et al. (2020) establish an approach based on semantic anchor points to assess the competitive environment. Lee et al., (2009) concentrated their research on how businesses might identify new business prospects based on their technical skills. In today’s competitive business environment, firms must make efforts to keep consumers pleased and maintain a niche in the market. Park et al. (2015) examined the connection between market value of corporations and their technology strategy in an investigation of the patenting of Korean companies. They came to the conclusion that because Korean companies are concentrating their technological diversification strategy on their core technology diversity, they can expect to see improvements in their performance in a short amount of time. Empirical findings of the review of literature under firm strategy confirm that firms seeking to have more competitive edge over its rivals should employ a wide-ranger technology diversification strategy when looking for new company prospects.

Cluster 2 emphasizes the critical role of competitive strategies, M&A, and collaborative innovation in firms’ competitiveness. It underscores the need for advanced patent analysis techniques and frameworks to guide decision-making in the pursuit of technological excellence. The findings suggest that firms aiming to gain a competitive edge should adopt a comprehensive technology diversification strategy and leverage their core technological competencies for sustainable growth and success in dynamic markets.

Cluster 3: Emerging Technology Forecasting and Evaluation

Cluster 3, focusing on emerging technology forecasting and evaluation, provides valuable insights into the dynamic landscape of technological trends and their implications. Firstly, it underscores the pivotal role of patents as indicators for detecting and forecasting technological trends. The emergence of carbon fiber reinforcing technology, as highlighted by Moehle and Caferoglu (2019), exemplifies how advancements in one area can have ripple effects across various industries, such as aviation, automobiles, bicycles, and wind turbines. Moreover, this cluster emphasizes the increasing interdependence and linkages between science and technology. Publications by Li et al. (2019), Wu et al. (2021), Lu et al. (2020), and Forestal et al. (2022) use citation network analysis to uncover trends and progress in specific research or technological fields. Li et al. (2019), for example, show how to use a framework that combines text mining and expert judgment to find the paths that technologies have taken over time and to guess how they will develop in the near future, focusing on perovskite solar cell technology. This approach is deemed crucial for informing R&D strategies.

A subset of studies within this cluster, such as those by Ena et al. (2016) and Miao et al. (2020), explores innovative pathways and future trends using semantic analysis and the Technology Road Mapping (TRM) technique. Miao et al. (2020), for instance, leverages Technology-Relationship-Technology (TRT) semantic analysis to extract TRT structures and dimensions of TRM, providing valuable insights into innovation pathways and trends. Furthermore, nanotechnology emerges as a significant theme in this cluster, with studies by Igami (2008) and Zhou et al. (2019) exploring applications of new materials and devices. These studies utilize patent mapping to examine nanotechnology development. Guo et al. (2012) employ multi-database NEST search results to develop algorithms for extracting technological components, major actors, and potential applications in the nanotechnology domain.

Ena et al. (2016) add to the literature by developing a new data clustering method for tracking technological developments. Recent research by Lu et al. (2020) proposes a novel method for identifying upcoming technologies by combining data mining methods with deep learning to overcome difficulties caused by insufficient training samples.

To sum up, the third cluster emphasizes the value of patents as leading indicators for monitoring technical developments. It highlights the need for novel approaches to forecasting short-term tendencies in technological development by integrating text mining, expert opinion, and data resources such as scientific publications and patents. Researchers, decision-makers, and technology professionals can gain helpful knowledge from this cluster as it reveals the growing interplay between science and technology and delves into emerging fields like nanotechnology.

Cluster 4: Patent Technical Intelligence

Cluster 4, centered on patent technical intelligence, sheds light on the critical role of patent analysis in converting patent information into valuable technical, commercial, and legal insights. Several notable findings and methodologies emerge from this cluster.

Firstly, the significance of patent keyword networks is underscored, with Choi et al. (2013) conducting trend analysis to identify how keywords influence network changes over time. Chen (2017) explores the relationship between technological information in patents and their supporting citations, proposing a deep learning-based technique for extracting meaningful insights from patents. A key takeaway from this cluster is the ability of patent analysis to define technology content through keyword correlations. Patents are recognized as a reliable source of technology intelligence. Scholars such as Park et al. (2012) and Chen et al. (2020) employ SAO-based semantic technological similarity to identify technology hotspots. They extract the SAO structure, incorporate
domain dictionaries and professional corpora, and employ Word2Vec for technology demand clustering. Furthermore, Souza et al. (2020) focus on evaluating abstractive and extractive summarization algorithms’ performance in generating terms directly related to patent claims. Zhang et al. (2022) discuss the challenges of representing the five stages of the evolutionary process—knowledge production, growth, obsolescence, transfer, and intergrowth—using a single term based on continuous time frequency. An et al. (2018) introduce a method to derive technology intelligence from patents, overcoming limitations of keyword-based network analysis. They demonstrate its potential using electric car patents as an example. Choi et al. (2022) present the PatentNet dataset, which records technical citation contexts based on textual data, metadata, and examiner citation data for a vast number of patents. This dataset supports technology planning decisions and has been utilized in intelligent patent tools, including deep learning techniques employed by researchers such as Lee et al. (2013) and Marx et al. (2022) to forecast new technological concepts.

In conclusion, Cluster 4 emphasizes the multifaceted nature of patent technical intelligence and its function in gaining valuable insights from patents. Research on innovation, innovation policy, and technology strategy all benefit greatly from up-to-date and reliable data. It is common knowledge that in the modern business world, patent intelligence is an indispensable tool for gaining an edge over the competition.

Cluster 5: Technological Convergence and Open-Endedness

The fifth grouping includes two related but separate ideas that help drive innovation and economic development: technological convergence and technical open-endedness. The first subset of this group is known as “technological convergence,” and it is defined by the merging of previously separate technological areas in order to create novel synergies. And Kim Sohn et al. (2020) describe an automated learning-based system for finding convergence trends. It does this by combining semantic data analysis with tried-and-true methods such as link forecasting and bibliometric evaluation. The research of novelty inside patent documents is still in its infancy, as noted by Walcott et al. (2017), and requires time-consuming manual techniques. Several studies highlight the significance of cross-industry technology convergence. Qin et al. (2021) introduce a method for evaluating cognitive proximity through mining patent description text using the LDA topic model, addressing the challenge of quantifying knowledge technologies. Cho et al. (2021) present a new paradigm for predicting technology convergence tendencies in different industries, contributing to technical and economic growth by anticipating emerging technology areas. Giordano et al. (2021) explore technological convergence using a novel methodology that combines text mining and dynamic network models based on Defense Patent Data. Lee et al. (2021) employs machine learning to predict multi-technology convergence, exemplified by a case study on pharmacological, bio-affecting, and body-treating components technology.

The second division of cluster 5 focuses on technological open-endedness, emphasizing the endless possibilities for innovation. Moehrle (2010) highlights the importance of measuring textual patent similarity for key patent administration tasks such as prior art analysis, infringement analysis, and patent mapping. Joo and Kim (2010) introduce a multi-dimensional contingency table representation of technological field co-occurrence and a relatedness metric to quantify the interrelatedness of technical fields, using Korean patent data for comparison. They provide a framework for tracking and predicting combinative innovation in science, technology, and innovation (ST&I). In summary, cluster 5 underscores the significance of both technological convergence, which drives innovation by bridging technology domains, and technological open-endedness, which highlights the endless possibilities for innovation across various fields. These themes contribute to economic growth and have practical applications in patent administration, prediction of technology trends, and creative idea generation.

CONCLUSIONS

The paper examined text mining analysis of patents in innovation studies. The study of the papers’ manual clustering leads us to the conclusion that conducting a systematic review is a scientifically rigorous method that helps one to discover and synthesize the available information addressing crucial concerns. The analysis of patent text analysis and innovation studies yields significant insights into the evolving landscape of research in this domain. Notably, this review identifies two journals, “Scientometrics” and “Technological Forecasting & Social Change,” as the primary repositories for publications in this field, underscoring their pivotal role in disseminating knowledge. The analysis also identifies five new and developing areas of study in this field, including the discovery of technological opportunities, the development of competitive strategies for businesses, the anticipation of technological developments, the analysis of patent data, and the study of technological convergence. There is a broad variety of applications for text mining techniques in patent analysis, and they are all showcased by the various clusters. In addition, the study draws attention to the methodological variety present in patent text analysis, with researchers using a broad range of approaches such as SAO-based semantic analysis, citation analysis, network analysis, and more. Among the many valuable insights that may be gained from patent data is the ability to identify technological components

https://journals.e-palli.com/home/index.php/ajebi
and predict future trends in technology. The analysis highlights the industrial focus of these articles, from tech companies to photovoltaics to business strategies to medicines and beyond, and it also suggests future research directions in the fields of robotics and medicine. Finally, this in-depth examination highlights the ever-changing nature of innovation studies and the crucial role of text mining approaches in revealing technical trends, possibilities, and strategies, laying a solid groundwork for future research and development in this area.

This in-depth look at patent text analysis and innovation studies shows how text mining methods are becoming more important for understanding technical changes, futures, and strategies. Numerous researches focus provide avenues for further study and development as well as valuable insights for experts in the domains of innovation and technology management.

**Future Research Directions**

It is possible to identify numerous prospective future research topics based on the examination of patent text analysis and innovation studies. To begin with, many of the evaluated studies in this research centered on specific companies or sectors, and the application of their proposed frameworks is confined to a single industry or sector (Li et al., 2019). Research on the use of text mining methods in patent research across different sectors is a promising area for future investigation. This might lead to a better understanding of the dynamics of innovation in different fields by shedding light on transferable best practices and industry-specific peculiarities. Second, while the evaluated studies showed that text mining techniques were employed for a variety of goals, including detecting technical components and trend prediction, they just scratched the surface of text summarization and classification. The efficient and organized examination of patent documents might be greatly aided by more study into the creation and use of text summary and classification systems. Thirdly, when technology develops further, it is crucial for studies of patent text analysis to adapt to new directions in the field. Natural language processing (NLP) and machine learning are two promising emerging technologies that might be used into future research to improve the quality and productivity of patent analysis. Research on how new technologies may affect innovation and patent management is a promising area.

Given the increasing relevance of patent analysis across several sectors, more research may investigate the moral and legal implications of text mining techniques in the context of patent data. Intellectual property, privacy, and ethical technology use are all topics that need to be investigated. Investigating how patent text analysis aligns with legal frameworks and ethical standards will be crucial as these technologies continue to evolve. Lastly, collaboration between researchers in the fields of text mining, innovation studies, and other relevant disciplines can lead to more comprehensive and impactful research outcomes. Future research agendas should emphasize interdisciplinary collaboration to address complex innovation challenges. This approach can foster a holistic understanding of how text mining techniques can be integrated into broader innovation and business strategies, transcending the boundaries of individual academic domains.

It is also observed from the review that none of the selected 162 articles employed text summarization and text categorization as a technique. To improve the accuracy of technological features, future studies should focus more on multi-source data fusion algorithms linked to technology commerce and technology demand. Therefore, conducting a study on text mining in innovation studies using text summarization and text categorization is a future research opportunity. Important contributions from other sources, such as books or discoveries under patent and innovation studies, with vital information may have gone unnoticed, skewing the findings of our systematic review because the study focused exclusively on published articles. Because the discipline of innovation studies is still evolving, future research should broaden the scope of the search beyond only published papers in journals. The majority of the approaches proposed in the articles are focused on a specific industry.

Finally, the applicability of these proposed frameworks is limited to a specific industry or sector (Li et al., 2019). Future studies should develop these approaches and do a complete study on a cross-industry analysis of the survey’s proposed approaches and develop. This will put the frameworks’ potency and strength to the test. Similarly, in terms of their proposed framework, the majority of research failed to investigate the real relationship between theory and practice. As such, we encourage further studies to develop this area.

**REFERENCES**


https://journals.e-palli.com/home/index.php/ajebi
Technological forecasting and social change, 176, 121431.


https://journals.e-palli.com/home/index.php/ajebi