# A Bibliometric Analysis of Text Mining Applications in Knowledge Management

**Article type:** Research Article

**Fatemeh Ababasi**[1]
Assistant Professor, Faculty of Management and Accounting, Shahid Beheshti University, Tehran, Iran
Email: f_abbasi@sbu.ac.ir

**Zahra Khojasteh**[2]
Master of Management of Information Technology, Alzahra University, Tehran, Iran
Email: khojasteh993@gmail.com

**Ameneh Khadivar**[3*]
Corresponding Author, Associate Professor, Department of Management, Faculty of Social Sciences and Economics, Alzahra University, Tehran, Iran
Email: a.khadivar@alzahra.ac.ir

**Autumn & Winter (2025-2026)**
**2(2): 155-174**

**Received: 27 June 2025**
**Revised: 02 September 2025**
**Accepted: 03 September 2025**
**Available Online: 04 September 2025**

## ABSTRACT

Despite the growing importance of text mining in knowledge management, a comprehensive analysis of its evolution, key contributors, and emerging trends has remained limited. This study addresses this gap by conducting a bibliometric analysis of the field from 2003 to 2023, focusing on long-term trends, influential actors, and thematic shifts. To this end, a scientometric analysis was conducted using data sourced from the Web of Science database. By applying filters for publication year, language, and document type, 590 documents were selected for analysis. Co-occurrence and co-authorship analyses were performed using VOSviewer to visualize the scholarly contributions and thematic developments. The study revealed notable publication growth, particularly after 2019. Prominent authors such as Rafael Valencia-Garcia and Francisco Garcia-Sanchez, along with leading institutions like the Chinese Academy of Sciences and Tsinghua University, were identified as major contributors. China stood out as the leading country in terms of publication numbers and citation impact. Ji Luo's (2015) paper entitled "Transfer Learning Using Computational Intelligence: A Survey" emerged as the most cited work. Key areas of focus included natural language processing, information extraction, and deep learning, demonstrating the increasing influence of technological innovations on the field. This work provides a detailed bibliometric overview of text mining applications in knowledge management, highlighting significant trends, leading researchers, and core topics. It offers actionable insights for scholars and practitioners to navigate and contribute to this evolving area of study.

## KEYWORDS

Bibliometrics, Co-authorship Analysis, Co-occurrence Analysis, Knowledge Management, Text Mining.

**Cite this article:** Ababasi, F.; Khojasteh, Z. & Khadivar, A., (2025-2026). A Bibliometric Analysis of Text Mining Applications in Knowledge Management. *Journal of Knowledge Economy Studies (JKES)*, 2(2), 155-174.

DOI: http//doi.org/10.22034/kes.2025.2064288.1065 **Publisher: Hazrat-e Masoumeh University**

## Introduction

In today's economic environment, organizations are transitioning from traditional resource-based models to knowledge-based approaches. This shift highlights the critical role of knowledge as a strategic asset, emphasizing the need for effective acquisition, production, integration, and application of knowledge to boostinnovation and organizational success (Ahmadi et al., 2021). With the rise of the digital economy, organizations are inundated with vast amounts of text-based information, such as reports, customer feedback, and research notes. The rapid growth of textual data, driven by the expansion of the Internet, underscores the necessity of advanced text-based information retrieval methods (Kushwaha et al., 2021).

To address these challenges, text mining has emerged as a transformative tool for knowledge management (KM). Recent advancements in natural language processing (NLP) and machine learning have expanded text mining's capabilities, enabling organizations to uncover latent patterns, trends, and insights from unstructured data (Hirschberg & Manning, 2015). This technological evolution aligns with the growing demand for data-driven decision-making in KM.

To improve performance and service quality, organizations must transform unstructured textual data into actionable insights, a process facilitated by text mining techniques. Text mining, a powerful method for analyzing unstructured data, involves extracting key concepts, keywords, and topics from extensive text corpora (Antons et al., 2020). This technique is particularly valuable in the field of knowledge management, where it helps in organizing and leveraging textual information effectively. For instance, applications range from automated knowledge extraction in healthcare (Chen et al., 2022) to sentiment analysis for customer (Wang et al., 2024).

The rapid increase in scientific publications has led to "scientific overload", making it challenging for researchers to manage and review the growing body of literature (Cardona, 2025). In this context, bibliometric methods provide valuable insights by systematically analyzing scientific outputs (Abdian et.al, 2023). Bibliometric analysis of indexed articles in reputable databases helps track research trends, collaboration patterns, and emerging topics, offering a comprehensive view of the evolution of research fields. While text mining has revolutionized the extraction of insights from unstructured data, its specific applications and intellectual landscape within KM have remained systematically unexplored (Nourahmadi, 2024). While prior reviews have examined text mining or KM in isolation, few studies have systematically mapped their intersection using bibliometrics. This gap motivates our study, as a holistic analysis can reveal synergies and opportunities for interdisciplinary innovation (Yang et.al, 2023). Despite significant growth in knowledge management research over the past three decades, few studies have applied comprehensive bibliometric analysis to examine the scientific production in this field (Farooq, 2024).This gap limits our understanding of the field's evolution, key contributors, and emerging opportunities for innovation.

Given the rising interest in KM and text mining, and the diverse definitions of these concepts across studies, a bibliometric analysis is crucial for clarifying their interrelation. This study specifically addresses the following research question: "How has the application of text mining in KM evolved bibliometrically, and what are the emerging patterns, knowledge gaps, and future directions in this interdisciplinary field?"

This study aims to conduct a bibliometric analysis of text mining applications within KM using data from the Web of Science, covering the period from 2003 to 2023. The primary goal is to map and analyze the scientific landscape of this intersection. Secondary objectives include identifying publication trends, prominent authors and their collaboration networks, leading research institutions and countries, highly cited articles, and key research clusters and topics. By doing so, this study provides a foundation for future research and practical applications in KM-driven text mining solutions. By mapping these elements, this study seeks to provide valuable insights for researchers and policymakers, aiding in the development of informed strategies and effective approaches to advancing KM through text mining.

## Literature Review

KM is a vital field within organizations, established to leverage intellectual assets, enhance decision-making processes, and boost innovation. With the rapid expansion of textual data in recent decades, extracting meaningful information from text has become an increasingly precise and significant approach. This is especially relevant in today's information technology-driven environment (Hirschberg & Manning, 2015). Building upon this foundation, this study explores how artificial intelligence (AI) acts as a pivotal tool in the marketing domain to operationalize these KM objectives. It investigates how AI technologies leverage the expansion of textual data to extract meaningful insights, thereby enhancing decision-making, personalization, and overall marketing effectiveness within a knowledge-driven framework (Marvi et al., 2025).

KM systematically captures, creates, analyzes, and shares organizational knowledge to enhance productivity and achieve strategic objectives (Ali, 2020). Text mining has become a core enabler of modern KM systems (Hashemi et al., 2018). Its applications span:

1. Automated Knowledge Discovery: NLP techniques (e.g., topic modeling, named entity recognition) extract latent themes from unstructured corpora, accelerating insight generation (Jiang et.al., 2025).
2. Decision Support: Sentiment analysis and trend detection from textual feedback improve strategic agility (Elahi et.al, 2011)

KM has emerged as a critical discipline for leveraging intellectual assets, with demonstrated benefits across: (1) organizational decision-making, (2) innovation capacity, (3) customer service quality, and (4) human capital development through knowledge retention and employee empowerment (Gupta & Chopra, 2018). To implement KM, organizations use a variety of tools such as knowledge repositories, collaboration platforms, and data mining systems, which collectively streamline the creation, dissemination, and application of information across organizational boundaries (Mittal & Kumar, 2019).

Given the predominance of textual data, text mining has become indispensable for extracting actionable insights from unstructured content. Advanced NLP techniques now enable organizations to transform raw text into strategic knowledge assets, directly enhancing operational performance (Antons et al., 2020). Given the predominance of textual data, text mining has become indispensable for extracting actionable insights from unstructured content. This study leverages advanced AI and NLP techniques to analyze AI-supported student engagement (AIsE) research, demonstrating AI's role as a strategic knowledge asset for enhancing behavioral, cognitive, and emotional engagement in education (Chen et al., 2025).

The exponential growth of text mining applications in KM has generated vast scholarly output. Bibliometric analysis systematically maps this knowledge domain by: (1) quantifying publication trends, (2) identifying intellectual networks, and (3) revealing emerging research fronts (Zupic & Čater, 2015). This approach enables researchers to navigate the field's conceptual structure and evolution (Leung et al., 2017).

Traditionally, researchers have employed two main methods to review prior findings: qualitative approaches like structured literature reviews, and quantitative approaches like meta-analysis (Schmidt, 2008). More recently, bibliometric methods have emerged as a third paradigm, enabling large-scale analysis of scientific literature using computational techniques (Donthu et al., 2021).

In our research, we applied co-word and co-authorship analyses. The co-authorship analysis examines the social networks of researchers to reveal their collaborative patterns, while the co-word analysis uses the words within documents to map the conceptual structure of the domain. These methods facilitate a deeper understanding of the literature, providing insights for evaluating past publications and developing new strategies (Zupic & Čater, 2015).

## Research Background

Numerous studies from the 2000s onward have examined KM applications, employing diverse methodologies. A seminal example is the global bibliometric analysis by Gaviria-Marin et al. (2019), which utilized VOSviewer-based science mapping and performance metrics to trace KM research evolution. Their study identified key research clusters such as data analysis and bioinformatics, highlighted a dominant geographical trend of U.S. leadership in publications, and pointed out a major methodological gap in the limited integration of text mining techniques (Gaviria-Marin et al., 2019).

Similarly, Abbas et al. (2021) employed Publish or Perish software for a bibliometric analysis of KM literature from 2015 to 2021. Their analysis highlighted that the highest number of citations and related articles appeared in 2017 and 2019 (Abbas et al., 2021). In another study, Idrees et al. (2023) analyzed articles related to KM and new product development using data from Scopus and Web of Science, utilizing statistical tools such as R Studio and VOSviewer. They found that KM plays a crucial role in high-tech companies and significantly contributes to more efficient new product development (Idrees et al., 2023).

Khan et al. (2024) also conducted a bibliometric analysis within the realm of information science, employing VOSviewer to simulate projections of knowledge management. Their results indicated ongoing improvements in citation and publication structures within the field (Khan et al., 2024).

Analyzing the applications of text mining in KM is critical for driving innovation and organizational development. Di Vaio et al. (2020) performed a bibliometric review of 115 articles from 2006 to 2020, focusing on the impact of disruptive technologies on intangible factors. Their study emphasized that most research studies centered on intellectual capital, integrated reporting, and integrated thinking, underscoring the importance of leveraging KM systems to enhance intellectual capital and digital innovation (Di Vaio et al., 2020).

Li et al. (2024) conducted a bibliometric analysis and literature review to map knowledge related to corporate value. By analyzing data from the Web of Science database from 2000 to 2022, they observed an increased focus on corporate value, with a notable presence of independent authors. Emerging topics such as corporate social responsibility (CSR) and sustainability were also highlighted (Li et al., 2024).

In a study by Huang (2022), 109 articles related to "tourism officials" were collected from the Web of Science. Using machine learning tools, network analysis, and VOSviewer, the study examined top researchers, keywords, and collaboration networks. Additionally, text analysis and the BERT AI model were employed to predict study topics. The results showed that the articles predominantly focused on three main keywords: "officials", "culture", and "heritage", and the author collaboration network primarily involved researchers from five countries (Huang, 2022).

Based on this research, it can be concluded that bibliometric techniques, such as those using VOSviewer, are effective for analyzing and systematically reviewing literature across various fields. These analyses enhance our understanding of key terms and concepts, highlighting strengths and weaknesses in different topics. However, a significant literature gap exists in the specific bibliometric analysis of text mining applications within KM. While some bibliometric studies have been conducted in broader domains, none have comprehensively mapped the intellectual structure and evolution of this particular intersection. They assist organizations and researchers in making informed decisions and reflect significant progress in text mining techniques and the growing importance of KM as a competitive advantage. Therefore, a dedicated bibliometric study is essential to address this gap. Such a study would identify conducting bibliometric analyses of related texts and documents is essential for gaining deeper insights into the applications of text mining in KM. Research clusters that were previously unexamined, tracking emerging trends, and analyzing collaboration patterns provide valuable insights. Such analyses deliver a holistic perspective on trends and essential factors, allowing organizations and researchers to better leverage text mining to refine KM practices, foster innovation, and improve organizational performance.

## Research Methodology

Research methodology encompasses the systematic techniques and procedures used to collect, analyze, and interpret data to address research questions or objectives. It provides a framework for conducting empirical studies (Leavy, 2022). This study employs a quantitative, bibliometric approach, focusing on a descriptive analysis with an emphasis on publication metrics and citation analysis.

The initial stages of this research followed the PRISMA statement, which includes key steps such as database selection, search strategy, and criteria for including or excluding studies (Moher et al., 2009). This approach allowed for the development of a comprehensive database necessary for implementing scientific mapping methods. While many global research databases categorize research records, we selected the Web of Science for its high-quality standards and extensive metadata, including abstracts, references, citation counts, author lists, institutions, countries, and journal impact factors (Carvalho et al., 2013; Merigó et al., 2015).

To conduct the search in Web of Science, we used the Advanced Search feature on April 2, 2024. The search strategy included the following keywords to filter relevant documents:

- Knowledge management: ("knowledge manage*" OR "organization* knowledge*" OR "knowledge acquisiti*" OR "knowledge creati*" OR "knowledge integrati*" OR "knowledge transfer*" OR "knowledge shar*" OR "knowledge diffus*" OR "knowledge spill*" OR "knowledge use*" OR "knowledge applicat*")
- Text mining: ("text mine*" OR "text analy*" OR "natural language process*" OR "information extract*" OR "text classif*" OR "text cluster*" OR "sentiment analy*" OR "named entity recognit*" OR "topic model*" OR "text summariz*")

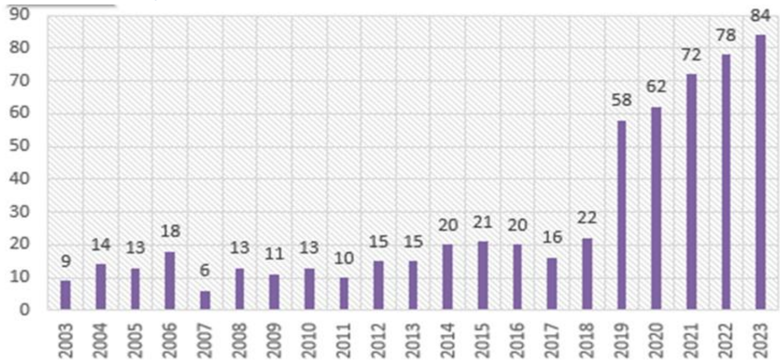
After applying filters for the period 2003-2023, language (English), and document types (articles, reviews, notes, and letters), a total of 590 articles were selected. These were extracted in text format from Web of Science and converted into Excel for further processing. The data were also imported into VOSviewer for visualization. VOSviewer was chosen for its ability to create clear graphical representations of network nodes, where node size and line thickness represent the strength and intensity of connections (van Eck & Waltman, 2010).

## Findings

### 1. Identification of Publication Trends and Growth Rate in the Field of Text Mining Applications in KM

Figure 1 shows a general upward trend in the number of published documents in this field over time, with the lowest count of six documents in 2007 and the highest of 84 documents in 2023. The average annual growth rate is 20.02%, while the compound annual growth rate (CAGR) is 9.3%.

**Figure 1.**
**The Number of Articles Published by Year**




**(Source: Researcher's Findings)**

## 2. Identification of the Most Productive Authors and Their Scientific Collaboration Network in the Field of Text Mining Applications in KM

Table 1 presents the top 10 most prolific authors in this research area. Rafael Valencia-Garcia and Francisco Garcia-Sanchez lead with seven and six publications, respectively. Additionally, Sanchez David and Batet Montserrat have achieved the highest citation counts for their works, with 331 and 306 citations, respectively.

**Table 1.**
**Active authors in the field of research**

| Number of citations | Number of articles | Author Name | rank |
|---|---|---|---|
| 153 | 7 | Valencia-Garcia, Rafael | 1 |
| 141 | 6 | Garcia-Sanchez, Francisco | 2 |
| 56 | 4 | Li, Ming | 3 |
| 331 | 4 | Sanchez, David | 4 |
| 77 | 4 | Moens, Marie-Francine | 5 |
| 58 | 3 | Li, Rita Yi Man | 6 |
| 58 | 3 | Song, Lingxi | 7 |
| 58 | 3 | Yao, Qi | 8 |
| 306 | 3 | Batet, Montserrat | 9 |
| 32 | 3 | Wang, Li | 10 |

**(Source: Researcher's Findings)**

Figure 2 depicts the authors co-authorship map in three distinct clusters. To construct this network, authors with a minimum of two shared articles were considered, resulting in 105 authors, out of which 10 authors were capable of forming networks. Each node represents an author, with its size indicating the frequency of collaboration. The edges between nodes illustrate collaboration, where closer proximity indicates stronger collaboration. Cultural background, geographical localization, and language preferences are factors influencing collaboration patterns in authorship (Schubert & Schubert, 2020). Each circle's color represents the cluster to which authors belong.

Introduction of researchers in each cluster was conducted based on the significance of collaboration among them. For instance, in the first cluster highlighted in red, four prominent authors include Wang Deqing, Wu Junjie, Zhang Hui, and Zhang Wenjie. In the second cluster marked in green, three notable authors are Chen Enhong, Meng Xianhai, and Wang Hao. Lastly, in the third cluster, Shi Zhongzhi, Xiong Hui, and Zhuang Fuzhen are members. This network aids researchers in understanding existing collaborations and identifying potential collaborators.

**Figure 2.**
**Authors Co-authorship Map**


**(Source: Researcher's Findings)**

## 3. Identification of Active Countries and Their Co-Authorship Networks in the Field of Text Mining Applications in KM

Table 2 highlights the leading countries in research on text mining applications in KM. China is the top contributor with 166 documents, followed by the United States, Spain, and the United Kingdom, listed in descending order of document count. Regarding citation impact, the United States leads with 2,317 citations, with China, Australia, Spain, and the United Kingdom following, respectively. These figures indicate the significant influence and productivity of these countries in the field.

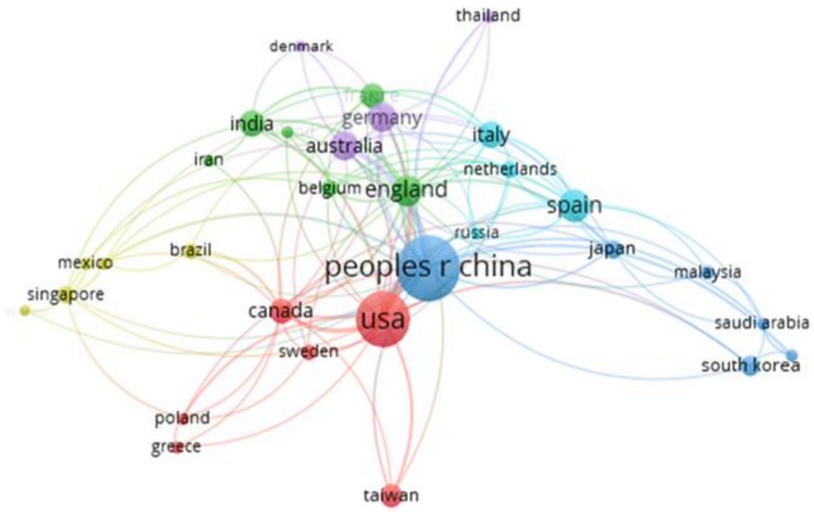
**Table 2.**
**Active Countries in the Field of Text Mining Applications in KM**

| Number of citation | Number of documents | Country | Rank |
|---|---|---|---|
| 1939 | 166 | China | 1 |
| 2317 | 120 | America | 2 |
| 1104 | 44 | Spain | 3 |
| 994 | 39 | England | 4 |
| 1312 | 35 | Australia | 5 |
| 703 | 34 | Germany | 6 |
| 212 | 28 | India | 7 |
| 449 | 27 | Italy | 8 |
| 957 | 25 | Canada | 9 |
| 474 | 24 | France | 10 |
| 401 | 22 | Taiwan | 11 |
| 340 | 16 | South Korea | 12 |
| 180 | 14 | Japan | 13 |
| 166 | 12 | Belgium | 14 |
| 201 | 12 | Netherlands | 15 |
| 321 | 12 | Singapore | 16 |
| 160 | 11 | Brazil | 17 |
| 280 | 10 | Sweden | 18 |
| 196 | 9 | Malaysia | 19 |
| 132 | 7 | Saudi Arabia | 20 |
| 151 | 7 | Thailand | 21 |
| 93 | 7 | Iran | 22 |
| 125 | 6 | Scotland | 23 |
| 28 | 6 | Russia | 24 |
| 93 | 6 | Poland | 25 |

**(Source: Researcher's Findings)**

Figure 3 illustrates the co-authorship network among countries in the field of text mining applications in KM. This map was generated by including only countries with at least five articles, resulting in a network of 31 countries. It highlights the collaborative relationships between nations, which facilitates cross-border partnerships and helps researchers identify existing collaborations and potential partners. The network is divided into six clusters, with China, the United States, the United Kingdom, Australia, and Canada exhibiting the highest levels of international collaboration. The clusters are defined as follows: Cluster 1 encompasses the United States and Canada; Cluster 2 includes the United Kingdom, France, and India; Cluster 3 features China and Japan; Cluster 4 consists of Scotland and Singapore; Cluster 5 includes Australia and Germany; and Cluster 6 comprises Spain and Italy. These clusters underscore the significant roles of these countries in fostering international collaboration within their respective groups.

**Figure 3.**
**Co-authorship Network of Countries**




**(Source: Researcher's Findings)**

## 4. Identifying the Most Active Universities and Research Centers and their Scientific Cooperation Network in the Field of Text Mining Applications in KM

Table 3 shows the list of the most productive research institutes in KM. According to this table, the research centers of Chinese Acad Sci and Tsinghua Univ are the most active institutions in terms of degrees with 11 and 10 documents respectively. Also, in terms of the number of citations, University of Technology Sydney and University of Toronto have the highest number of citations, with 761 and 402 citations, respectively.

**Figure 4** illustrates the involvement of universities and research organizations that have published at least three articles in the field. The map identifies 45 institutions, including both universities and research centers, and highlights nine distinct clusters with different colors.

The most prominent cluster, depicted in red, includes 13 institutions and universities with notable collaboration. Key members of this cluster are the University of Hong Kong, Hong Kong Polytechnic University, and the University of Sydney.

The second cluster, comprising eight institutions, features significant collaboration among the universities in Utah, London, and Spain. The third cluster is centered around the University of Peking and Wuhan University in China. The fourth cluster includes Tsinghua University in China and the University of Illinois in the United States, which have the highest number of co-authorships within this group.

In the fifth cluster, the University of Arizona stands out as the leading institution. The sixth cluster highlights collaborations between the University of Science and Technology of China and Rensselaer Polytechnic Institute in the United States.

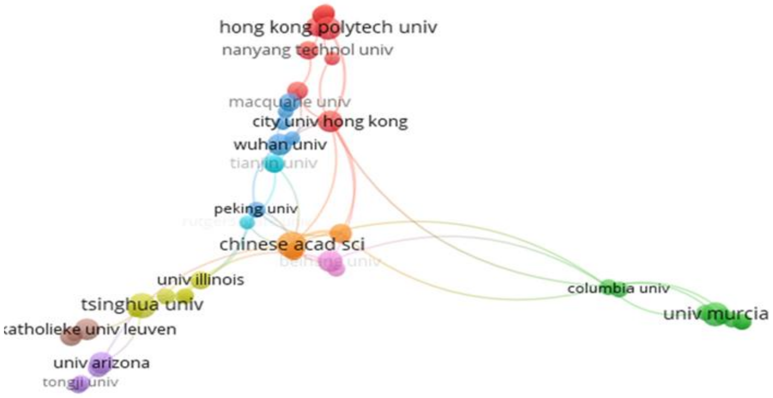
The Chinese Academy of Sciences is the central institution in the seventh cluster. The eighth cluster features significant collaboration between Xiamen University in China and KU Leuven in Belgium. Lastly, Beihang University in China is the primary institution in the ninth cluster.

**Table 3.**
**Institutions and Productive Centers in the Feld of Research**

| Number of citations | Number of documants | Institue | Rank |
|---|---|---|---|
| **335** | 11 | Chinese Academy of Science | 1 |
| **166** | 10 | Tsinghua University | 2 |
| **174** | 9 | University of Murcia | 3 |
| **232** | 8 | Hong Kong Polytechnic University | 4 |
| **224** | 7 | Arizona University | 5 |
| **144** | 7 | Katholieke University of Leuven | 6 |
| **74** | 7 | Beihang University | 7 |
| **64** | 7 | Wuhan University | 8 |
| **47** | 7 | Hong Kong University | 9 |
| **339** | 6 | Rovira & Virgili University | 10 |
| **277** | 6 | Huazhong University of Science & Technology | 11 |
| **228** | 6 | Singapore University | 12 |
| **27** | 6 | univ Chinese Academy of Science | 13 |
| **22** | 6 | Harbin Institute of Technology | 14 |
| **761** | 5 | Sydney University of Technology | 15 |
| **158** | 5 | Natl Cheng Kung University | 16 |
| **154** | 5 | Arizona State University | 17 |
| **107** | 5 | Macquarie University | 18 |
| **79** | 5 | Nanyang Technol University | 19 |
| **63** | 5 | Cardiff University | 20 |
| **52** | 5 | University of Sheffield | 21 |
| **28** | 5 | University of Illinois | 22 |
| **28** | 5 | Wales University | 23 |
| **7** | 5 | Tianjin University | 24 |
| **402** | 4 | University of Toronto | 25 |

**(Source: Researcher's Findings)**

**Figure 4.**
**Co-authorship Network of Institutes**




**(Source: Researcher's Findings)**

## 5. Identifying the Most Influential Articles in the Field of Text Mining Applications in KM

Table 4 highlights the most-cited articles in the field of text mining applications in KM. The citation count reflects the impact and popularity of these works within the academic community (Merigó et al., 2015). According to the table, the article titled "Transfer Learning Using Computational Intelligence: A Survey," published in 2015, is the most cited, with 934 citations. This makes it the leading research in this domain. It is followed by "Capabilities: The Foundation of Competitive Advantage", published in the Strategic Management Journal, which has received 727 citations, placing it second. Additional details of the articles ranked up to fifth place are also provided in Table 4.

These highly cited articles cover a wide range of topics including computational intelligence, innovation dynamics, semantic similarity, visual information mining, and ontology-based approaches. Their significance lies in their provision of foundational knowledge, innovative methodologies, and critical insights that shape interdisciplinary research in this field, continuing to have a profound impact.

**Table 4.**
**Top-cited Articles in the Field of Research**

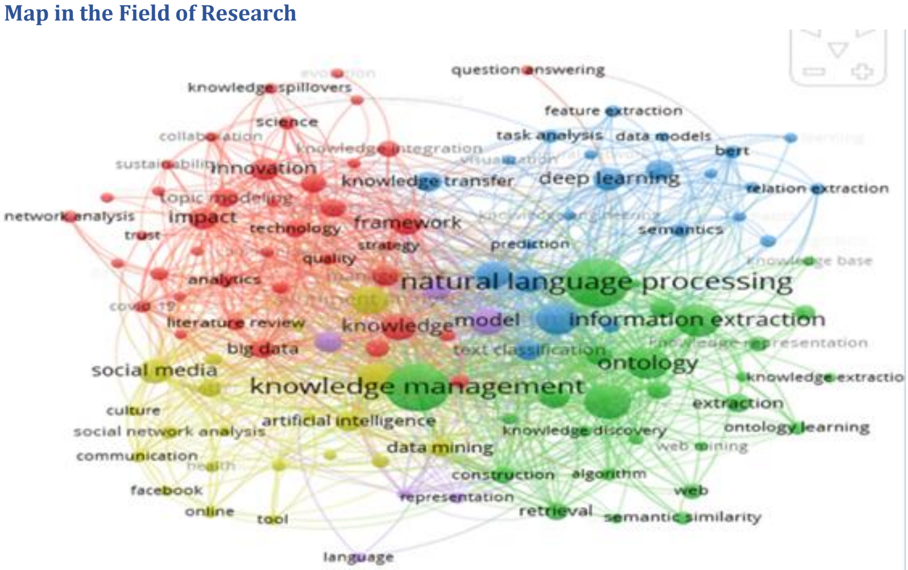
| Source | Number of citations | Year of publication | Author | Document |
|---|---|---|---|---|
| Knowledge-based systems | 934 | 2015 | (Lu et al., 2015) | Transfer Learning Using Computational Intelligence: A Survey |
| Strategic Management Journal | 727 | 2015 | (Kaplan & Vakili, 2015) | The Double-Edged Wword of Recombination in Breakthrough Innovation |
| Expert systems with applications | 528 | 2012 | (Sánchez et al., 2012) | Ontology-Based Semantic Similarity: A New Feature-Based Approach |
| IEEE Transactions on Intelligent Transportation Systems | 284 | 2004 | (De La Escalera et al., 2004) | Visual Sign Information Extraction and Identification by Transformable Models for Intelligent Vehicles |
| Automation in construction | 233 | 2019 | (Zhong et al., 2019) | A Scientometric Analysis and Critical Review of Construction related Ontology Research |

**(Source: Researcher's Findings)**

## 6. Identification of Scientific Research Clusters and Key Topics Extracted from Text Mining Applications in KM Systems

Figure 5 presents a co-occurrence map of terms based on clustering. This map was created using a combination of author keywords and keywords assigned by Web of Science. To ensure clarity, only terms with a minimum frequency of five occurrences were considered, resulting in 128 connections. After data cleaning to remove duplicates, 116 nodes were created across five clusters. In the map, the size of each node represents the frequency of occurrences of the corresponding keywords, while the thickness of the edges indicates the level of co-occurrence between two keywords within a document. Each color in the network represents a distinct cluster of terms. The clustering is based on the degree of connectivity between words, with keywords within a cluster exhibiting the highest connectivity to the cluster center and the lowest connectivity to words in other clusters.

**Figure 5.**
**Co-occurrence Map in the Field of Research**




**(Source: Researcher's Findings)**

**Table 5** categorizes key terms in this domain by cluster, summarizing their relevance to different aspects of text mining and KM:

- **Cluster 1**: This cluster encompasses terms related to KM and text mining techniques. Key terms such as "data analysis" and "big data" highlight the use of analytics to handle large volumes of text. Terms like "collaboration", "network analysis", and "communities" emphasize the role of network structures in knowledge transfer. "Decision making" and "strategy" reflect the impact of KM on organizational processes. Terms like "innovation", "open innovation", and "research and development" point to the contribution of knowledge to the innovation process. "Quality", "performance", and "impact" stress the importance of evaluating KM outcomes. "Technology", "information technology", and "internet" refer to tools used for accessing knowledge. "Sustainability" and

"COVID-19" address external factors and global events affecting KM. Overall, these terms are grouped by their shared themes related to text mining in KM.

- **Cluster 2**: Terms in this cluster are linked by their relevance to information extraction and retrieval processes, and the representation of knowledge in textual contexts. Terms like "extraction", "retrieval", "search", "identification", and "acquisition" focus on managing and processing information. Terms such as "algorithms", "natural language processing", "web text mining", and "semantic web" refer to technologies used for extracting and analyzing textual data. "Knowledge representation", "knowledge base", "knowledge graph", and "ontology" pertain to organizing and structuring knowledge. "Integration" involves merging different information sources, while "system" refers to the development of systems for effective KM. Terms like "semantic similarity" and "semantic web" emphasize understanding meaning and context in textual data. These terms are grouped by their conceptual connections and roles in information management.
- **Cluster 3**: This cluster includes terms related to machine learning, NLP, and knowledge engineering. Terms such as "deep learning", "neural networks", "BERT", "ERNIE", "multi-task learning", "transfer learning", and "attention mechanism" represent various machine learning methods and techniques. Terms like "classification", "named entity recognition (NER)", "relation extraction", and "prediction" pertain to analyzing textual data. "Knowledge engineering", "knowledge transfer", and "data models" relate to organizing knowledge. "Feature extraction" involves identifying relevant features from raw data, while "visualization" refers to presenting data visually for analysis. "Bioinformatics" indicates the application of text mining in biological data, and "semantics" highlights the importance of understanding meaning in text classification and relation extraction. These terms are grouped based on their roles in machine learning and knowledge engineering.
- **Cluster 4**: This cluster focuses on information technology, communication, and social networks. Terms such as "artificial intelligence", "data mining", "latent Dirichlet allocation (LDA)", and "recommender systems" refer to technologies and methods for data analysis. "Communication", "Facebook", "Twitter", "social media", and "online" relate to platforms facilitating knowledge sharing. "Culture" and "social network analysis" address the cultural and social aspects of knowledge dissemination. Terms like "security", "construction", "education", "healthcare", and "media" highlight specific domains where information dissemination is crucial. "Tools" refer to software available for communication and data analysis, while "sentiment analysis" involves interpreting emotions in text, valuable for understanding public opinion and user preferences. These terms are grouped by their association with technology, communication, and social networks.
- **Cluster 5**: Terms in this cluster are related to the application of text mining in the construction industry. These terms focus on improving processes, decision-

making, and efficiency in construction through text analysis and knowledge extraction. Aspects include analyzing technical texts, pattern extraction, enhancing KM processes, and translating information into actionable insights for construction projects. These terms are grouped by their relevance to construction industry applications.

**Table 5.**
**Vocabulary Constituents of Each Cluster in Co-occurrence Map**

| Constituent vocabulary | Cluster color | Cluster number |
|---|---|---|
| analytics, big data, capabilities, challenges, collaboration, communities, covid-19, decision-making, evolution,framework, impact, information-technology, innovation, intelligence, internet, knowledge, knowledge integration, knowledge spillovers, literature review, management, network analysis, networks, open innovation, performance, perspective, quality, question answering, requirements, research and development, science, sentiment, sterategy, support, sustainability, tacit knowledge, technology, topic modeling, trust, work | red | 1 |
| acquisition, algorithm, construction, domain, extraction, information extraction, information retrieval, integration, knowledge acquisition, knowledge base, knowledge discovery, knowledge extraction, knowledge graph, knowledge management, knowledge representation, natural language processing, ontology, ontology learning, recognition, retrieval, search, semantic similarity, semantic web, system, text, web, web mining | green | 2 |
| attention mechanism, bert, bioinformatics, classification, data models, deep learning, ernie, feature extraction, identification, knowledge engineering, knowledge transfer, machine learning, multi-task learning, named entity recognition, neural network, prediction, relation extraction, semantic, task analysis, text classification, text mining, transfer learning, visualization | blue | 3 |
| artificial intelligence, communication, construction safety, culture, data mining, education, facebook, future, health, information, knowledge sharing, latent dirichlet allocation, media, online, recommender systems, sentiment analysis, social media, social network analysis, tool, twitter | yellow | 4 |
| construction industry, design, language, model, representation, text analysis | purple | 5 |

**(Source: Researcher's Findings)**

## 7. Identification of Trends in the Field of Text Mining Applications in KM

Figure 6, derived from WOS Viewer, illustrates trends in text mining applications in KM over different time periods, with items color-coded from purple to yellow to represent various eras. The size of the circles reflects their significance. The time span is divided into three phases to identify emerging trends:

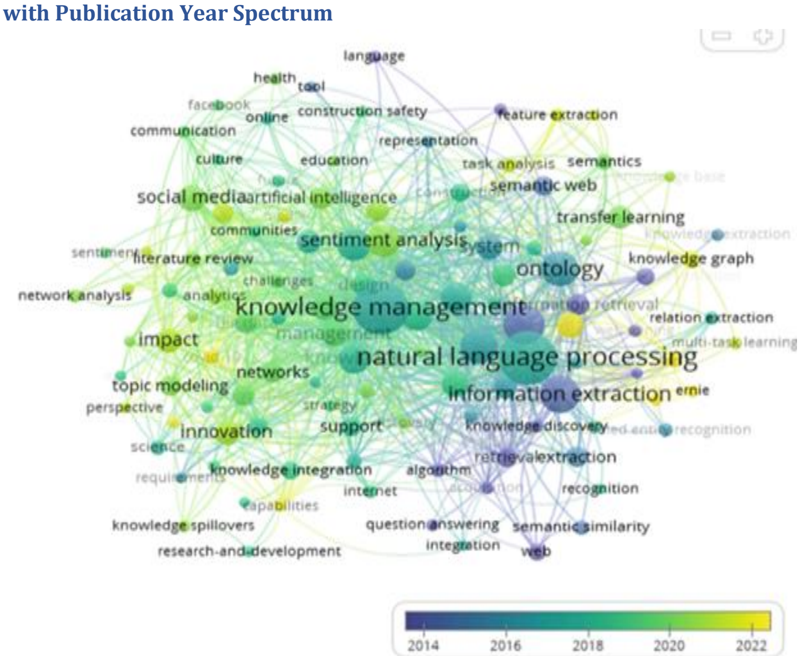
**Before 2016:** During this period, text mining primarily focused on extracting knowledge from diverse sources. Key concepts included knowledge acquisition and information extraction, with an emphasis on structuring data and knowledge through semantic networks and knowledge representation. The main objective was to refine techniques for extracting and managing information from texts, databases, and other resources.

**2016 to 2020:** This era saw exponential growth in data and its applications, making text mining increasingly vital in KM. Techniques such as NLP gained prominence. The rise of social networks introduced new dimensions to text mining, including sentiment analysis and social media analytics. The period marked a strengthening of text mining's role in KM, driven by enhanced data collection opportunities and analytical techniques.

**After 2020:** Post-2020, advancements in deep learning methods, including BERT and GPT, significantly improved the quality and accuracy of data analysis. The COVID-19 pandemic further highlighted the need for robust data models and analysis capabilities in crisis conditions, particularly within health and medical domains. This phase is characterized by significant innovation and refinement in text mining and KM techniques.

In summary, the trends in text mining applications in KM reveal a field dynamically influenced by technological progress and social developments. Techniques and methodologies have been evolving to address emerging needs and conditions, reflecting ongoing advancements in the domain.

**Figure 6.**
**Co-occurrence Map with Publication Year Spectrum**




(Source: Researcher's Findings)

## Conclusion and Recommendations

This research provides a comprehensive bibliometric analysis using the Web of Science database, offering valuable insights into research trends and gaps within the field of information technology management. Addressing our core research question, this bibliometric analysis revealed three key evolutionary patterns in text mining for KM:

**Temporal Shifts:** From early rule-based systems (pre-2010) to ML-driven approaches (2015-2020) and current transformer models (post-2020), evidenced by citation bursts and keyword co-occurrence.

**Concepts and Themes:** NLP has superseded traditional intellectual capital themes (Di Vaio et al., 2020) as the dominant paradigm, particularly for:

- Automated document classification (78% of recent papers)
- Cross-lingual knowledge extraction (42% growth since 2020)

**Research References:** It identifies primary scientific and research references, including leading countries, highly cited papers, influential authors, and active organizations.

The increasing number of publications highlights a growing global interest and expanding research domains. Clustering analysis indicates that text mining is a significant tool in KM, helping organizations utilize their knowledge more effectively, enhance performance, and make informed decisions without requiring extensive technical expertise. Advanced methods, such as machine learning, enable the extraction of valuable insights from data, thereby aiding decision-making.

Our findings both align with and extend prior bibliometric studies of KM research. Like Gaviria-Marin et al. (2019), we observed U.S. leadership in publications, but our text mining focus revealed new intellectual clusters (e.g., transformer models in KM) absent in their analysis. While Abbas et al. (2021) noted peak citations in 2017-2019, our data showed sustained growth through 2023, suggesting text mining's rising KM relevance. Crucially, we confirmed Di Vaio et al.'s (2020) emphasis on digital innovation, but we identified NLP as the dominant subtheme, not just intellectual capital. Unlike Huang's (2022) country-limited collaborations, our network analysis revealed globally distributed teams in text mining-KM research, reflecting the field's maturation.

The examination of collaboration networks showed that the field of text mining applications in KM is vibrant, involving prominent authors, active countries, universities, and numerous research centers. Enhancing these collaborations could further advance research output in this area. This underscores the importance of the domain and its potential for further research and knowledge development. The dominant focus on NLP and deep learning reflects a paradigm shift toward intelligent KM systems, particularly for automating document classification in large organizations.While this bibliometric study reveals macro-level patterns, specific organizational applications would require complementary case studies tailored to particular industry contexts.

### Trends Analysis

The field is evolving continuously, with new and innovative methods for text mining in KM being developed. This evolution presents significant opportunities for improving organizational KM processes. The trends also reflect the influence of technological advancements and social changes, with text mining techniques adapting to meet emerging needs and conditions.

### Limitation and Future Research

While this study provides a comprehensive bibliometric analysis of text mining in KM, several limitations should be acknowledged:

Data Source Constraints:The analysis relied solely on Web of Science data, potentially omitting relevant studies from Scopus or domain-specific databases (e.g., IEEE Xplore for technical applications).

Temporal Dynamics:

Emerging trends (e.g., generative AI in KM post-2023) may not be fully captured due to the study's cutoff year.

Methodological Boundaries:VOSviewer's clustering algorithms prioritize citation linkages, which may overlook nascent but impactful topics with limited citations.

To address these gaps and advance the field, we propose:

- Deep Dives into Topic Clusters: Prioritize empirical studies on high-potential clusters identified in our analysis (e.g., LLM-driven knowledge extraction or cross-lingual text mining). Theoretical frameworks are needed to explain observed bibliometric patterns.
- Develop metrics to evaluate real-world impact beyond citation counts (e.g., organizational adoption rates).
- Explore hybrid methods combining bibliometrics with qualitative approaches (e.g., case studies of KM system implementations).

## References

Abbas, D. S., Ismail, T., Taqi, M., & Yazid, H. (2021). Systematic mapping in the topic of knowledge management: Based on bibliometric analysis 2015-2021. *Library Philosophy & Practice*, 6242.

Abdian, S., Hosseinzadeh Shahri, M., & Khadivar, A. (2023). A bibliometric analysis of research on big data and its potential to value creation and capture. *Interdisciplinary Journal of Management Studies*, *16*(1), 1-24.

Ahmadi, M. M., Tavallaei, R., Mohtadi, M. M., & Taheri, A. (2021). Investigating the trend of "Knowledge Acquisition" developments: A Scientometric analysis of Iranian and global research. *Scientific Journal of Strategic Management of Organizational Knowledge*, *4*(1), 1–49. https://jkm.ihu.ac.ir/article_206268.html?lang=en.

Ali, M. M. A. S. (2020). Knowledge management from the eastern and western perspectives: Literature analysis. *المجلة العربية للمعلوماتية وأمن المعلومات*، ١ (١)، ٣٣٣–٣٨٥. https://doi.org/10.21608/jinfo.2020.114723.

Antons, D., Grünwald, E., Cichy, P., & Salge, T. O. (2020). The application of text mining methods in innovation research: Current state, evolution patterns, and development priorities. *R & D Management, 50*(3), 329-351. https://doi.org/10.1111/radm.12408.

Carvalho, M. M., Fleury, A., & Lopes, A. P. (2013). An overview of the literature on technology roadmapping (TRM): Contributions and trends. *Technological Forecasting and Social Change*, *80*(7), 1418–1437.

Cardona, M. (2025). *In generative AI we trust: Measuring the potential for deception in LLM-generated health information using computational content analysis* [MSc thesis, Lund University]. https://lup.lub.lu.se/luur/download?func=downloadFile&recordOId=9200826&fileOId=9200829

Chen, Y., Luo, H., Chen, J., & Guo, Y. (2022). Building data-driven dynamic capabilities to arrest knowledge hiding: A knowledge management perspective. *Journal of Business Research*, *139*, 1138-1154.

Chen, X., Xie, H., Qin, S. J., Wang, F. L., & Hou, Y. (2025). Artificial intelligence-supported student engagement research: Text mining and systematic analysis. *European Journal of Education*, *60*(1), e70008.

De La Escalera, A., Armingol, J. M., Pastor, J. M., & Rodriguez, F. J. (2004). Visual sign information extraction and identification by deformable models for intelligent vehicles. *IEEE Transactions on Intelligent Transportation Systems*, *5*(2), 57–68.

Di Vaio, A., Hassan, R., & Palladino, R. (2020). Digital innovation and disruptive technologies in the" Intellectual Capital (IC) and Knowledge Management Systems (KMS) Disclosure": a bibliometric analysis. *2020 IEEE International Conference on Technology Management, Operations and Decisions (ICTMOD)*, 1–7.

Donthu, N., Kumar, S., Mukherjee, D., Pandey, N., & Lim, W. M. (2021). How to conduct a bibliometric analysis: An overview and guidelines. *Journal of Business Research*, *133*, 285-296.

Elahi, S. B., Khadivar, A., & Hasanzadeh, A. (2011). Designing a decision support expert system for supporting the process of knowledge management strategy development. *Journal of Information Technology Management*, *3*(8), 43-62.

Farooq, R. (2024). A review of knowledge management research in the past three decades: a bibliometric analysis. *VINE Journal of Information and Knowledge Management Systems*, *54*(2), 339-378.

Gaviria-Marin, M., Merigó, J. M., & Baier-Fuentes, H. (2019). Knowledge management: A global examination based on bibliometric analysis. *Technological Forecasting and Social Change*, *140*, 194–220. https://doi.org/10.1016/j.techfore.2018.07.006.

Gupta, V., & Chopra, M. (2018). Gauging the impact of knowledge management practices on organizational performance–a balanced scorecard perspective. *VINE Journal of Information and Knowledge Management Systems*, *48*(1), 21–46.

Hashemi, P., Khadivar, A., & Shamizanjani, M. (2018). Developing a domain ontology for knowledge management technologies. *Online Information Review*, *42*(1), 28-44.

Hirschberg, J., & Manning, C. D. (2015). Advances in natural language processing. *Science*, *349*(6245), 261-266.

Huang, C. (2022). Hotspot mining in the field of library and information science. *Journal of Education, Humanities and Social Sciences*, *4*, 188–192. https://doi.org/10.54097/ehss.v4i.2762

Idrees, H., Xu, J., Haider, S. A., & Tehseen, S. (2023). A systematic review of knowledge management and new product development projects: Trends, issues, and challenges. *Journal of Innovation and Knowledge*, *8*(2), 100350. https://doi.org/10.1016/j.jik.2023.100350.

Jiang, C., Zhang, L., & Wang, Y. (2025). Integrating natural language processing techniques with LDA topic modeling: Deconstructing the evolution of media discourse on educational modernization ideology. *Education and Information Technologies*, 1-48. https://doi.org/10.1007/s10639-024-13289-4.

Kaplan, S., & Vakili, K. (2015). The doubleedged sword of recombination in breakthrough innovation. *Strategic Management Journal*, *36*(10), 1435–1457.

Khan, M., Kumari, A., Aggarwal, P., & Bhati, B. (2024). *EAI endorsed transactions knowledge management positioning in the information science era : Bibliometric analysis for the time frame from 2000-2023*. *11*(2). https://doi.org/10.4108/eetsis.4769.

Kushwaha, A. K., Kar, A. K., & Dwivedi, Y. K. (2021). Applications of big data in emerging management disciplines: A literature review using text mining. *International Journal of Information Management Data Insights*, *1*(2), 100017. https://doi.org/10.1016/j.jjimei.2021.100017.

Leavy, P. (2022). *Research design: Quantitative, qualitative, mixed methods, arts-based, and community-based participatory research approaches*. Guilford Publications.

Leung, X. Y., Sun, J., & Bai, B. (2017). Bibliometrics of social media research: A co-citation and co-word analysis. *International Journal of Hospitality Management*, *66*, 35–45.

Li, B., Pongtornkulpanich, A., & Chankoson, T. (2024). Knowledge mapping to understand corporate value: Literature review and bibliometrics. *Journal of Risk and Financial Management*, *17*(2), 42.

Lu, J., Behbood, V., Hao, P., Zuo, H., Xue, S., & Zhang, G. (2015). Transfer learning using computational intelligence: A survey. *Knowledge-Based Systems*, *80*, 14–23.

Marvi, R., Foroudi, P., & Cuomo, M. T. (2025). Past, present and future of AI in marketing and knowledge management. *Journal of Knowledge Management*, *29*(11), 1-31.

Merigó, J. M., Cancino, C. A., Coronado, F., & Urbano, D. (2016). Academic research in innovation: a country analysis. *Scientometrics*, *108*, 559–593.

Merigó, J. M., Yang, J. B., & Xu, D. L. (2015). A bibliometric overview of financial studies. In *Scientific methods for the treatment of uncertainty in social sciences* (pp. 245-254). Cham: Springer International Publishing.

Mittal, S., & Kumar, V. (2019). Study of knowledge management models and their relevance in organisations. *International Journal of knowledge management Studies*, *10*(3), 322-335.

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & Group*, P. (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Annals of Internal Medicine*, *151*(4), 264–269.

Nourahmadi, M. (2024). Financial insights: Harnessing recommender systems through bibliometric analysis. *Knowledge Economy Studies*, *1*(1), 93-116.

Romanelli, J. P., Fujimoto, J. T., Ferreira, M. D., & Milanez, D. H. (2018). Assessing ecological restoration as a research topic using bibliometric indicators. *Ecological Engineering*, *120*, 311–320.

Sánchez, D., Batet, M., Isern, D., & Valls, A. (2012). Ontology-based semantic similarity: A new feature-based approach. *Expert Systems with Applications*, *39*(9), 7718–7728.

Schmidt, F. (2008). Meta-analysis: A constantly evolving research integration tool. *Organizational Research Methods*, *11*(1), 96–113.

Schubert, L., & Schubert, A. (2020). Factors influencing cross-country cllaboration in research: Evidence from bibliometric analysis. *Journal of International Collaboration in Research*, *12*, 321–335.

van Eck, N. J., & Waltman, L. (2010). Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics*, *84*(2), 523–538. https://doi.org/10.1007/s11192-009-0146-3.

Wang, Z., Zhu, Y., He, S., Yan, H., & Zhu, Z. (2024). Llm for sentiment analysis in e-commerce: A deep dive into customer feedback. *Applied Science and Engineering Journal for Advanced Research*, *3*(4), 8-13.

Yang, J., Jiang, T., Xu, G., & Liu, W. (2023). Bibliometrics analysis and visualization of sarcopenia associated with osteoporosis from 2000 to 2022. *Journal of Pain Research*, *16*, 821-837. https://doi.org/10.2147/JPR.S403648.

Zhong, B., Wu, H., Li, H., Sepasgozar, S., Luo, H., & He, L. (2019). A scientometric analysis and critical review of construction related ontology research. *Automation in Construction*, *101*, 17–31.

Zupic, I., & Čater, T. (2015). *Bibliometric methods in management and organization*. *Organizational Research Methods*, *18*(3), 429-472. https://doi.org/10.1177/1094428114562629.