

A Bibliometric Study of Software Architecture

Loso Judijanto

IPOSS Jakarta, Indonesia and losojudijantobumn@gmail.com

ABSTRACT

This study provides a comprehensive bibliometric analysis of software architecture research to map its intellectual structure, collaboration patterns, and thematic evolution. Data were collected from the Scopus database, covering publications from 2000 to 2026. The analysis employs quantitative bibliometric techniques and network visualization using VOSviewer, including co-authorship, citation, keyword co-occurrence, overlay, and density analyses. The results indicate that software architecture research has experienced substantial growth, characterized by increasing global collaboration and the emergence of dominant research hubs such as China, the United States, and India. Citation analysis highlights that the most influential studies are largely driven by technological tools, frameworks, and artificial intelligence innovations, reflecting a shift toward application-oriented research. Furthermore, keyword analysis reveals that while traditional topics such as architectural design and system performance remain central, emerging themes such as machine learning, cloud computing, and interdisciplinary applications—particularly in bioinformatics—are gaining prominence. The overlay and density visualizations confirm a temporal transition toward more adaptive, efficient, and data-driven architectural paradigms. Overall, this study contributes to a deeper understanding of the development and current trends in software architecture research, offering insights into future directions and supporting strategic research planning in this rapidly evolving field.

Keywords: *Software Architecture, Bibliometric Analysis, Co-Authorship Network, Keyword Co-Occurrence, Research Trends*

1. INTRODUCTION

Software architecture has become a central pillar in the development of modern information systems, serving as the blueprint that defines system structure, components, interactions, and guiding design principles. As software systems have evolved from monolithic applications to highly distributed and scalable ecosystems, the role of architecture has expanded significantly [1], [2]. Contemporary paradigms such as microservices, cloud-native systems, and edge computing demand architectural decisions that not only ensure functional correctness but also address non-functional requirements such as scalability, reliability, maintainability, and security [3], [4]. Consequently, software architecture is no longer viewed merely as a technical concern, but as a strategic asset that influences organizational agility and long-term system sustainability.

Over the past two decades, research in software architecture has grown rapidly, reflecting both technological advancements and increasing system complexity. Early studies primarily focused on architectural styles and design patterns, providing foundational frameworks for structuring software systems [5], [6], [7]. However, the emergence of new computing paradigms—such as service-oriented architecture (SOA), cloud computing, and DevOps—has shifted scholarly attention toward dynamic, adaptive, and continuous architectural practices. More recently, the integration of artificial intelligence, Internet of Things (IoT), and cyber-physical systems has further broadened the scope of software architecture research, introducing new challenges related to interoperability, real-time processing, and system resilience [8], [9], [10].

Despite this rapid growth, the expanding body of literature has led to fragmentation, making it increasingly difficult to identify dominant themes, influential contributors, and emerging research directions. Researchers and practitioners often face challenges in navigating this vast knowledge

landscape, which spans multiple disciplines and application domains [11], [12]. Traditional literature reviews, while valuable, may be limited in their ability to capture large-scale patterns and relationships across extensive datasets. This highlights the need for systematic and quantitative approaches to synthesize existing knowledge and uncover the intellectual structure of the field.

Bibliometric analysis has emerged as a powerful method for addressing this challenge by enabling the quantitative evaluation of scientific publications [13], [14]. Through techniques such as co-authorship analysis, citation analysis, and keyword co-occurrence mapping, bibliometric methods allow researchers to identify collaboration networks, influential studies, and thematic trends within a research domain. Tools like VOSviewer have further enhanced the ability to visualize complex relationships in large datasets, making it possible to explore the evolution of research topics and the interconnectedness of scholarly contributions. In the context of software architecture, bibliometric analysis offers a systematic way to map its development, providing insights that are difficult to obtain through conventional review approaches.

Although several studies have examined specific aspects of software architecture, such as design patterns or cloud-based systems, a comprehensive bibliometric mapping of the field remains limited. Existing reviews often focus on narrow subtopics or short time spans, thereby lacking a holistic perspective on the evolution of software architecture research. This gap underscores the importance of conducting a large-scale bibliometric study that captures the full breadth of the field over an extended period.

Therefore, this study aims to conduct a comprehensive bibliometric analysis of software architecture research using data retrieved from Scopus. By analyzing publications from 2000 to 2026, this study seeks to: (1) examine the patterns of scientific collaboration through co-authorship analysis; (2) identify the most influential publications and authors through citation analysis; and (3) explore the intellectual structure and emerging trends of the field through keyword co-occurrence analysis.

The findings of this study are expected to provide a structured and data-driven overview of software architecture research, offering valuable insights for academics, practitioners, and policymakers. By identifying key research clusters, influential contributors, and emerging topics, this study contributes to a deeper understanding of the field and supports the development of future research agendas in software architecture.

2. METHODS

This study employs a quantitative bibliometric approach to systematically analyze the development of software architecture research. Bibliometric analysis is widely used to evaluate scientific publications by identifying patterns in authorship, citations, and research themes. The data for this study were collected from the Scopus database, which is recognized for its extensive coverage of high-quality peer-reviewed literature [15], [16]. The search was conducted using relevant keywords such as “software architecture,” “software design,” and related terms, covering publications from 2000 to 2026. To ensure data quality and relevance, only journal articles, conference papers, and review papers written in English were included, while duplicates and unrelated documents were excluded through a screening process.

After data collection, the bibliographic information—including authors, titles, abstracts, keywords, citations, and affiliations—was exported in CSV format for further analysis. The study utilizes bibliometric mapping and network analysis techniques to explore the intellectual structure of the field. Visualization and network construction were performed using VOSviewer, which enables the identification of relationships among authors, publications, and keywords [17], [18].

Three main types of analysis were conducted: co-authorship analysis to examine collaboration patterns among researchers and institutions; citation analysis to identify influential publications and authors; and keyword co-occurrence analysis to uncover major research themes and emerging trends within software architecture.

To enhance the reliability of the findings, threshold criteria were applied during the analysis process, such as minimum numbers of documents, citations, or keyword occurrences required for inclusion in the network maps. The results were then interpreted using a descriptive and exploratory approach, focusing on identifying clusters, link strengths, and temporal developments in the field. This methodological framework allows for a comprehensive understanding of the evolution, structure, and dynamics of software architecture research, providing a solid foundation for drawing meaningful conclusions and identifying future research directions.

3. RESULT AND DISCUSSION

3.1 Author Collaboration Analysis

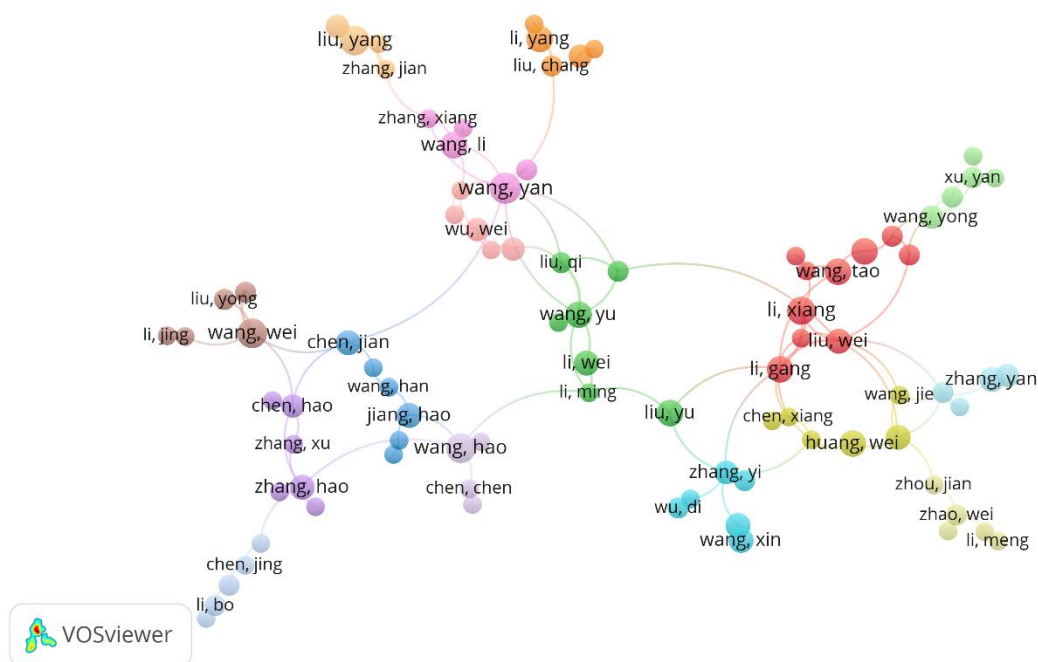


Figure 1. Author Visualization
Source: Data Analysis

Figure 1 presents the author co-authorship network, illustrating the structure of collaboration among researchers in software architecture studies. The visualization reveals several distinct clusters, each representing groups of authors who frequently collaborate within specific research subdomains. Notably, certain authors such as Wang, Li, and Liu appear as central nodes connecting multiple clusters, indicating their significant role as collaboration hubs and knowledge bridges across research communities. The density of connections within clusters suggests strong intra-group collaboration, while the inter-cluster links reflect cross-topic or interdisciplinary research interactions. However, the presence of smaller, loosely connected clusters also indicates that some research groups remain relatively isolated, potentially limiting knowledge diffusion. Overall, this network highlights a moderately interconnected but still somewhat fragmented collaboration landscape, suggesting opportunities to strengthen global and interdisciplinary partnerships in advancing software architecture research.

3.2 Country Contribution Analysis

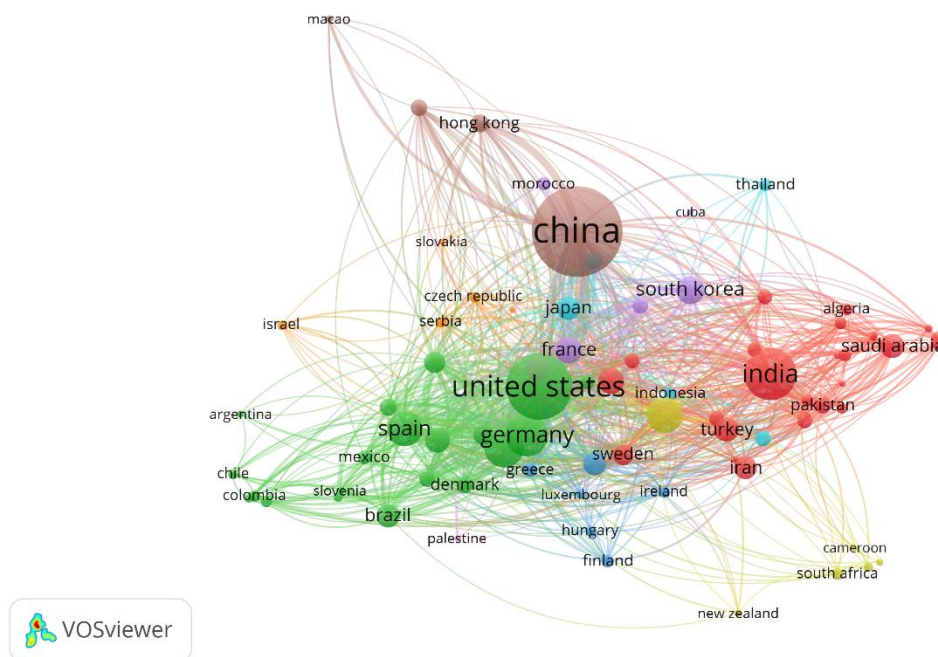


Figure 2. Country Visualization

Source: Data Analysis

Figure 2 illustrates the country-level collaboration network in software architecture research, highlighting the global distribution and interconnectedness of scholarly contributions. The visualization shows that countries such as China, the United States, and India occupy central and dominant positions, indicated by their larger node sizes and extensive linkages with other countries. These nations act as major hubs of research productivity and international collaboration. European countries such as Germany, Spain, and France also form strong collaborative clusters, demonstrating a well-integrated regional research network. Additionally, the presence of connections between developed and developing countries—such as collaborations involving Indonesia, Turkey, and Iran—indicates an increasing globalization of research activities. Despite this, some regions still appear less connected, suggesting disparities in research participation and collaboration intensity. Overall, the network reflects a highly interconnected global research landscape, with a few dominant countries driving knowledge production while also fostering international partnerships across regions.

3.3 Citation Analysis

Table 1. The Most Impactful Literatures

Citations	Authors and year	Title
8006	Giannozzi, P.,Andreussi, O.,Brumme, T.,... Wu, X.,Baroni, S. (2017)	Advanced capabilities for materials modelling with Quantum ESPRESSO
7791	Abramson, J.,Adler, J.,Dunger, J.,... Hassabis,	Accurate structure prediction of biomolecular interactions with AlphaFold 3

Citations	Authors and year	Title
	D.,Jumper, J.M. (2024)	
6902	Akiba, T.,Sano, S.,Yanase, T.,Ohta, T.,Koyama, M. (2019)	Optuna: A Next-generation Hyperparameter Optimization Framework
5504	Perez-Riverol, Y.,Csordas, A.,Bai, J.,... Brazma, A.,Vizcaíno, J.A. (2019)	The PRIDE database and related tools and resources in 2019: Improving support for quantification data
5307	Lee, J.,Yoon, W.,Kim, S.,... So, C.H.,Kang, J. (2020)	BioBERT: A pre-trained biomedical language representation model for biomedical text mining
4799	Rueden, C.T.,Schindelin, J.,Hiner, M.C.,... Arena, E.T.,Eliceiri, K.W. (2017)	ImageJ2: ImageJ for the next generation of scientific image data
3965	Jouppi, N.P.,Young, C.,Patil, N.,... Wilcox, E.,Yoon, D.H. (2017)	In-datacenter performance analysis of a tensor processing unit
3707	Choi, Y.,Choi, M.,Kim, M.,... Kim, S.,Choo, J. (2018)	StarGAN: Unified Generative Adversarial Networks for Multi-domain Image-to-Image Translation
2821	Di Renzo, M.,Zappone, A.,Debbah, M.,... De Rosny, J.,Tretjakov, S. (2020)	Smart Radio Environments Empowered by Reconfigurable Intelligent Surfaces: How It Works, State of Research, and the Road Ahead
2353	Kiranyaz, S.,Avci, O.,Abdeljaber, O.,... Gabbouj, M.,Inman, D.J. (2021)	1D convolutional neural networks and applications: A survey

Source: Scopus, 2026

Table 1 highlights the most impactful literatures in the dataset, as indicated by their exceptionally high citation counts, reflecting their foundational and cross-disciplinary influence. The dominance of works such as Giannozzi et al. (2017) and Abramson et al. (2024) demonstrates the strong impact of computational tools and artificial intelligence—particularly platforms like Quantum ESPRESSO and breakthroughs like AlphaFold—on scientific and engineering research. Similarly, studies such as Akiba et al. (2019) on Optuna and Lee et al. (2020) on BioBERT indicate the growing importance of machine learning frameworks and domain-specific AI models in advancing research efficiency and capability. Other highly cited works, including ImageJ2 (Rueden et al., 2017)

and tensor processing unit analysis (Jouppi et al., 2017), further emphasize the role of infrastructure and tools in enabling large-scale data processing and experimentation. Overall, the table suggests that the most influential studies are not limited to theoretical contributions but are largely driven by practical tools, platforms, and frameworks that significantly enhance research productivity across multiple domains, highlighting a shift toward technology-driven innovation in modern scientific ecosystems.

3.4 Keyword Co-Occurrence and Research Themes

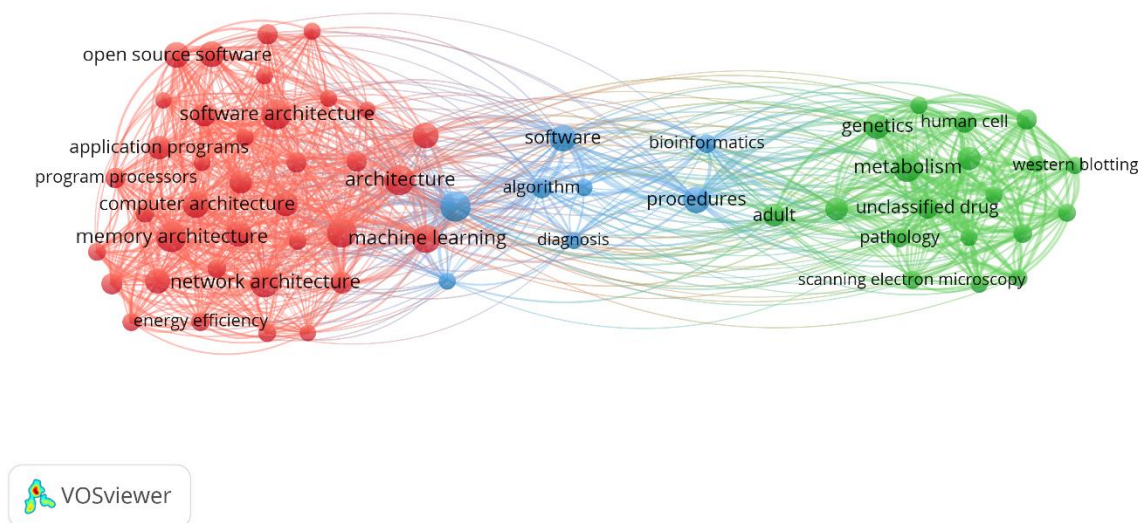


Figure 3. Network Visualization

Source: Data Analysis

Figure 3 presents the keyword co-occurrence network, revealing the intellectual structure and thematic clustering within software architecture research. The visualization shows three dominant clusters: the red cluster, which centers on core topics such as software architecture, computer architecture, network architecture, and energy efficiency, indicating the foundational and technical aspects of the field; the blue cluster, which includes terms like software, algorithm, and machine learning, reflecting the integration of computational intelligence and data-driven approaches; and the green cluster, which is oriented toward interdisciplinary applications, particularly in bioinformatics and biomedical domains such as genetics, metabolism, and pathology. The strong connections between these clusters suggest an increasing convergence between traditional software architecture and emerging application areas, especially those involving artificial intelligence and life sciences. This pattern indicates that software architecture is not only evolving within its own domain but is also becoming a critical enabler in cross-disciplinary innovation, highlighting a shift toward more application-driven and integrative research directions.

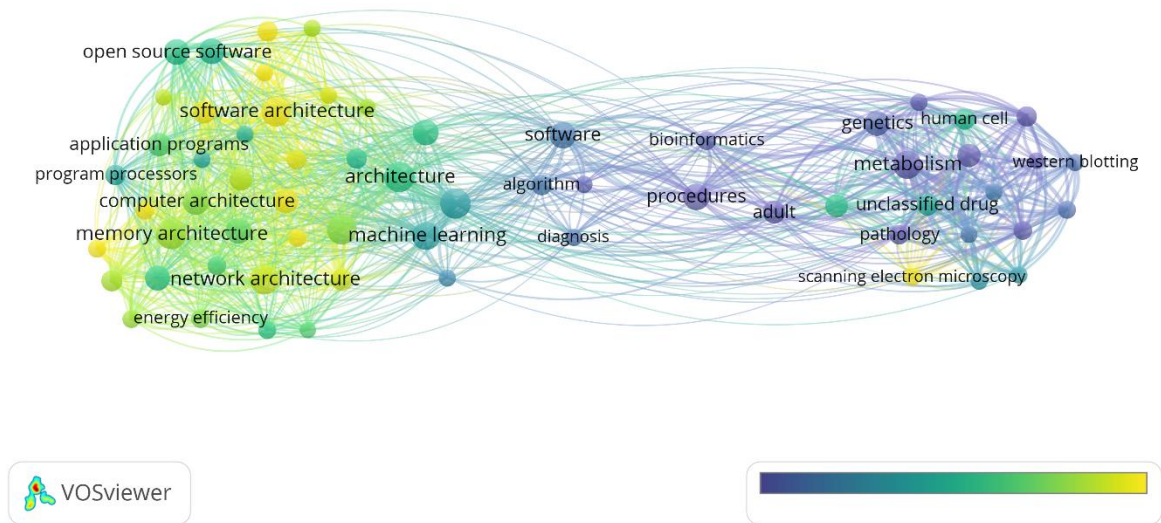


Figure 4. Overlay Visualization

Source: Data Analysis

Figure 4 presents the overlay visualization, which illustrates the temporal evolution of research topics in software architecture based on average publication year. The color gradient—from blue (earlier studies) to yellow (more recent studies)—indicates how research themes have developed over time. The visualization shows that earlier research (blue–purple areas) is concentrated in interdisciplinary domains such as bioinformatics, genetics, and biomedical applications, reflecting initial integrations of computational methods into life sciences. In contrast, more recent research trends (green–yellow areas) are strongly associated with core software architecture topics, including software architecture, network architecture, energy efficiency, and open source software. Additionally, the growing prominence of machine learning and algorithm-related keywords in the transitional color range suggests a shift toward intelligent and adaptive systems. Overall, this pattern indicates that while software architecture research initially expanded into applied domains, recent developments emphasize strengthening core architectural concepts while integrating modern technologies, highlighting a dynamic and continuously evolving research landscape.

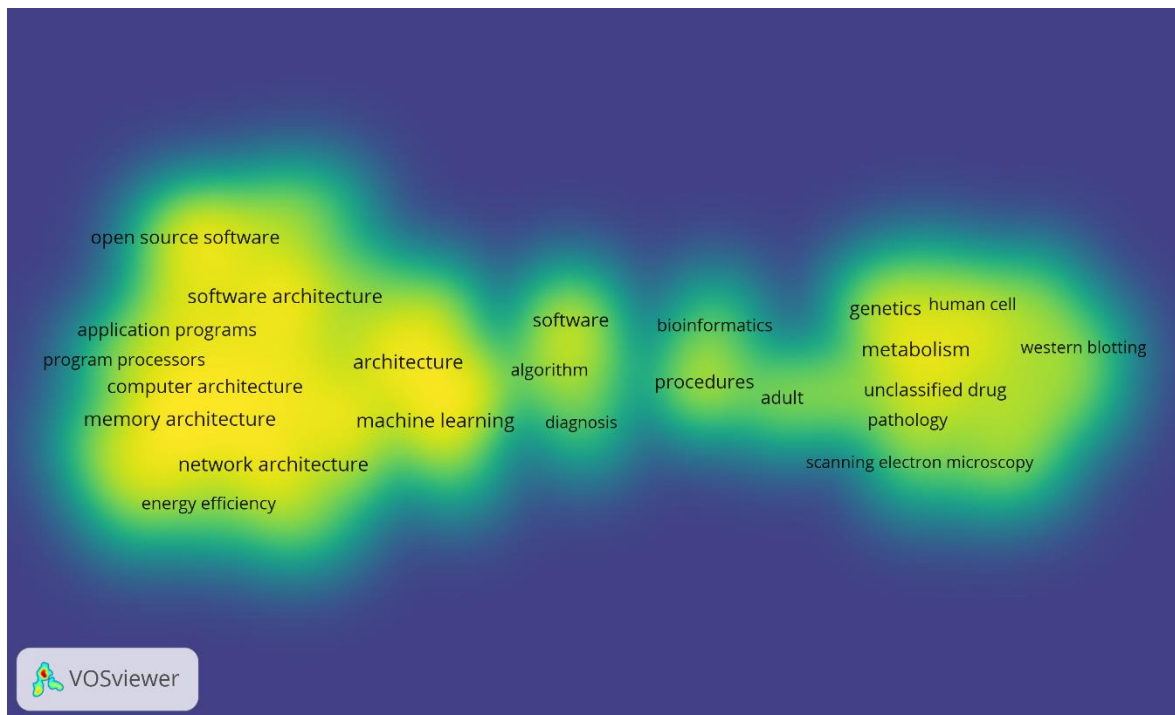


Figure 5. Density Visualization

Source: Data Analysis

Figure 5 presents the density visualization of keyword occurrences, highlighting the intensity and concentration of research topics within the software architecture domain. Areas with brighter yellow coloration indicate higher keyword frequency and stronger research focus, while darker regions represent less explored topics. The visualization shows a high-density concentration around core themes such as software architecture, architecture, machine learning, computer architecture, and network architecture, suggesting that these topics form the central backbone of the field. Additionally, a secondary but notable density appears in interdisciplinary areas such as bioinformatics, genetics, metabolism, and pathology, indicating the application of software architecture concepts in life sciences and biomedical research. The presence of connecting terms like software, algorithm, and procedures between these clusters reflects the role of computational methods as a bridge between core engineering and applied domains. Overall, this density pattern suggests that while foundational architectural topics remain dominant, there is a growing and sustained interest in cross-disciplinary applications, reinforcing the expanding relevance of software architecture in diverse scientific fields.

Discussion

The findings of this study reveal that software architecture research has evolved into a highly dynamic and interconnected field, characterized by strong collaboration networks and increasing global participation. The co-authorship analysis demonstrates that while several dominant researchers and clusters lead the field, the overall structure remains moderately fragmented. This suggests that although collaboration is well established, there are still opportunities to strengthen interdisciplinary and cross-regional partnerships. Enhancing such collaborations could accelerate knowledge diffusion and foster innovation, particularly in emerging areas that are currently represented by smaller and less connected research groups.

From the citation analysis, it is evident that the intellectual foundation [8], [12], [19] of software architecture is shaped not only by theoretical contributions but also by practical tools, frameworks, and technological innovations. Highly cited works tend to focus on enabling technologies such as machine learning frameworks, optimization tools, and computational

platforms, indicating a shift toward application-driven research. This trend reflects the increasing demand for scalable, efficient, and intelligent systems in both academia and industry. At the same time, the persistence of foundational studies highlights the importance of maintaining strong theoretical underpinnings while adapting to rapidly changing technological landscapes.

The keyword co-occurrence, overlay, and density visualizations collectively indicate a significant transformation in research themes over time. Traditional topics such as architectural design, system performance, and network architecture remain central, but they are increasingly complemented by emerging themes such as machine learning, cloud computing, and interdisciplinary applications in fields like bioinformatics. The overlay analysis further shows a temporal shift toward more recent topics related to efficiency, open-source ecosystems, and intelligent systems, suggesting that the field is moving toward more adaptive and data-driven architectures. This evolution underscores the growing complexity of modern software systems and the need for innovative architectural solutions that can accommodate diverse technological requirements.

Overall, the results suggest that software architecture research is transitioning toward a more integrative and application-oriented discipline. The convergence of multiple domains—including artificial intelligence, life sciences, and advanced computing—indicates that software architecture is becoming a critical enabler of cross-disciplinary innovation. Future research should focus on strengthening global collaboration, addressing the challenges of system complexity and scalability, and exploring new architectural paradigms that support emerging technologies. By doing so, the field can continue to advance in a way that is both theoretically robust and practically relevant in addressing real-world problems.

CONCLUSION

This study has systematically mapped the landscape of software architecture research using a bibliometric approach, revealing important insights into its structure, evolution, and future trajectory. The findings demonstrate that the field is characterized by strong global collaboration, with several dominant countries and research groups playing a central role in knowledge production. At the same time, the presence of smaller and less connected clusters suggests that opportunities remain to enhance interdisciplinary and international collaboration.

The analysis also shows that the intellectual foundation of the field is increasingly shaped by practical tools, computational frameworks, and emerging technologies such as artificial intelligence and machine learning, indicating a shift from purely theoretical discussions toward more application-driven and solution-oriented research. Furthermore, the evolution of key research themes highlights a transition toward more dynamic, scalable, and intelligent systems. Overall, software architecture research is becoming more integrative and interdisciplinary, bridging core engineering principles with diverse application domains, and future studies should focus on addressing system complexity, scalability, and interoperability while strengthening global collaboration.

REFERENCES

- [1] A. Kotb, M. Abdel-Kader, A. Allam, H. Halabi, and ..., "Information technology in the British and Irish undergraduate accounting degrees," *Account. ...*, 2019, doi: 10.1080/09639284.2019.1588135.
- [2] Y. Blount, B. Abedin, S. Vatanasakdakul, and ..., "Integrating enterprise resource planning (SAP) in the accounting curriculum: a systematic literature review and case study," *Account. ...*, 2016, doi: 10.1080/09639284.2016.1138136.
- [3] N. Al-Abdelmalek *et al.*, "Transforming Challenges into Opportunities for Qatar's Food Industry: Self-Sufficiency, Sustainability, and Global Food Trade Diversification," *Sustain.*, vol. 15, no. 7, 2023, doi: 10.3390/su15075755.
- [4] A. Azoulay, N. M. Doris, K. B. Filion, and ..., "The use of the transition cost accounting system in health services research," 2007, *resource-allocation.biomedcentral* doi: 10.1186/1478-7547-5-11.
- [5] T. Bauer, P. O. Antonino, and T. Kuhn, "Towards architecting digital twin-pervaded systems," *2019 IEEE/ACM 7th ...*, 2019, [Online]. Available: <https://ieeexplore.ieee.org/abstract/document/8882862/>
- [6] M. Al-Okaily, R. Alghazzawi, A. F. Alkhwalidi, and ..., "The effect of digital accounting systems on the decision-

- making quality in the banking industry sector: a mediated-moderated model," *Glob. Knowl. ...*, 2022, doi: 10.1108/GKMC-01-2022-0015.
- [7] A. Okon, N. Jagannath, I. Elgendi, J. M. H. Elmirghani, and ..., "Blockchain-enabled multi-operator small cell network for beyond 5G systems," ... *Netw.*, 2020, [Online]. Available: <https://ieeexplore.ieee.org/abstract/document/9136597/>
- [8] W. Liu, Y. Liang, X. Bao, J. Qin, and M. K. Lim, "China's logistics development trends in the post COVID-19 era," *Int. J. ...*, 2022, doi: 10.1080/13675567.2020.1837760.
- [9] L. A. Amaral, E. De Matos, R. T. Tiburski, F. Hessel, and ..., "Middleware technology for IoT systems: Challenges and perspectives toward 5G," ... *Things 5G ...*, 2016, doi: 10.1007/978-3-319-30913-2_15.
- [10] P. Jiang, Y. Van Fan, J. Zhou, M. Zheng, X. Liu, and ..., "Data-driven analytical framework for waste-dumping behaviour analysis to facilitate policy regulations," *Waste Manag.*, 2020, [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0956053X19307950>
- [11] I. P. Chochliouros, M. A. Kourtis, A. S. Spiliopoulou, and ..., "Energy efficiency concerns and trends in future 5G network infrastructures," *Energies*, 2021, [Online]. Available: <https://www.mdpi.com/1996-1073/14/17/5392>
- [12] M. A. Kashem, M. Shamsuddoha, and T. Nasir, "Smart Transportation and Carbon Emission from the Perspective of Artificial Intelligence, Internet of Things, and Blockchain: A Review for Sustainable Future," 2024.
- [13] H. Xie and T. C. Lau, "Evidence-Based Green Human Resource Management: A Systematic Literature Review," *Sustain.*, vol. 15, no. 14, 2023, doi: 10.3390/su151410941.
- [14] F. C. Fenerich, K. Guedes, N. H. M. Cordeiro, G. de S. Lima, and A. L. G. De Oliveira, "Energy efficiency in industrial environments: an updated review and a new research agenda," *Rev. Gestão e Secr. (Management Adm. Prof. Rev.)*, vol. 14, no. 3, pp. 3319–3347, 2023, doi: 10.7769/gesec.v14i3.1802.
- [15] L. Zhao, Y. Zhao, J. Du, A. Desloge, Z. Hu, and G. Cao, "Mapping the Research on Health Policy and Services in the Last Decade (2009–2018): A Bibliometric Analysis," *Front. Public Heal.*, vol. 10, p. 773668, 2022.
- [16] M. Akbari, M. Khodayari, A. Khaleghi, M. Danesh, and H. Padash, "Technological innovation research in the last six decades: a bibliometric analysis," *Eur. J. Innov. Manag.*, vol. 24, no. 5, pp. 1806–1831, 2021.
- [17] M. N. Ngoc, C. N. Thanh, and N. N. Van, "The impact of financial decisions and manager characteristics to firm's financial performance in SMEs: Bibliometric and systematic analysis," *Soc. Sci. Humanit. Open*, vol. 12, p. 101899, 2025.
- [18] M. Castillo-Vergara, I. Quispe-Fuentes, and J. Poblete, "Technological innovation in the food industry: A bibliometric analysis," *Eng. Econ.*, vol. 32, no. 3, pp. 197–209, 2021.
- [19] E. Hadjielias, M. Christofi, P. Christou, and ..., "Digitalization, agility, and customer value in tourism," ... *Forecast. Soc. ...*, 2022, [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0040162521007654>