Content Analysis of CITC Global Conference-Construction in the 21st Century- A Quantitative Look at the Text Corpus of CITC-1 to CITC-12

Farnaz Jafari[1], Muhammad Umer[1], Rana Muhammad Irfan Anwar[1], Salman Azhar, Ph.D.[1], Rizwan U. Farooqui, Ph.D.[2]

[1] Auburn University, Auburn AL 36849, USA
[2] Mississippi State University, Starkville MS 39762, USA
mzu0009@auburn.edu

Abstract
The construction industry is continually evolving, driven by technological advancements, environmental concerns, and socioeconomic factors. The Construction in the 21st Century (CITC) Global conferences provide a platform for industry leaders to discuss these challenges and opportunities. This paper presents a comprehensive content analysis focusing on the textual corpus of the CITC Global conferences from CITC-1 to CITC-12. This study identifies key trends and themes shaping the construction industry by utilizing the word cloud analysis, pareto charts, and topic modeling. The analysis reveals a predominant focus on project management, cost control, sustainable development, and the integration of advanced technologies such as Building Information Modeling (BIM). Additionally, the study highlights the industry's commitment to addressing these critical issues through systematic research and study. By understanding the primary themes and evolving priorities discussed in the CITC conferences, this paper offers valuable insights for researchers, practitioners, and policymakers navigating the dynamic landscape of the construction industry. The findings contribute to more effective and informed decision-making processes, paving the way for future research and strategic planning in the construction sector.

Keywords
Content Analysis, Construction, Word Clouds, Pareto Chart, Topic Modeling.

1. Introduction

The construction industry faces constant evolution, driven by technological advancements, environmental concerns, and socioeconomic factors. The Construction in the 21st Century (CITC) Global conferences serve as a valuable platform for industry leaders to discuss these challenges and opportunities. This paper delves into the content of CITC Global conferences including CITC-1 to CITC-12 held between 2002 and 2022, employing content analysis to identify key trends and themes shaping the construction industry. By analyzing conference proceedings abstracts, this study aims to:

- Identify the most prominent topics and sub-topics addressed within the conferences.
- Explore any shifts or continuities in focus areas over the chosen period.
- Gain insights into the evolving priorities and challenges faced by construction professionals.

Additionally, this paper examines the evolution of modern factors such as Building Information Modeling (BIM), which shows distinct trends over the analysis period, and highlights the growing interest in emerging. This analysis will contribute to a better understanding of current trends and future directions within the construction industry. It can also offer valuable insights for researchers, practitioners, and policymakers to navigate the ever-changing landscape of construction.

Content analysis in construction research involves systematically examining and interpreting the content of various construction-related documents and studies to extract valuable insights and trends. Studies like those by Mohamed and Tran (2022) and Zhang et al. (2021) highlight the use of content analysis to explore topics such as e-

inspection techniques in highway projects and knowledge extraction from construction documents using natural language processing. Additionally, Rashidi et al. (Rashidi et al., 2023) demonstrate how content analysis can be applied to investigate bidding trends in construction projects over several decades, providing a comprehensive overview of research contributions, keywords, and co-occurrence networks. By employing content analysis methodologies, researchers can gain a deeper understanding of construction practices, identify knowledge gaps, and pave the way for future studies in the field.

Content analysis employs various techniques such as word clouds, Pareto charts, and topic modeling to extract and visualize information from large datasets. Word clouds, which visually represent the frequency of words in a document by displaying more frequent words in larger, darker type, are useful for summarizing qualitative data and identifying key themes quickly (Cidell, 2010; Dicle & Dicle, 2018). Pareto charts, which graphically represent the most significant factors in a dataset, can help prioritize issues by showing that a small number of causes often account for a large portion of the problems, thus aiding in decision-making and process improvement (Jagannathan et al., 2022). Topic modeling, particularly using Latent Dirichlet Allocation (LDA), is a powerful method for identifying latent themes within large text corpora. For instance, in construction management, LDA has been used to detect trends such as BIM and lean construction, highlighting the evolving focus areas in the field (Bilge & Yaman, 2021; Liu et al., 2019). These methods collectively enhance the understanding of construction documents, from contract terms to social media posts, by extracting and visualizing knowledge that can inform policy and strategic decisions (Li et al., 2024; Zhang et al., 2021). The integration of these content analysis tools provides a comprehensive approach to managing and interpreting the vast amount of unstructured data generated in the construction industry, ultimately contributing to more effective and informed decision-making processes.

In this study content analysis tools including word cloud, Pareto chart, and topic modeling using LDA were utilized to examine twenty years of construction research presented at the CITC Global Conference.

2. Methodology

The methodology for analyzing the CITC Global Conference involves a comprehensive content analysis of the text corpus from CITC-1 to CITC-12 (Fig.1). This process includes three primary methods: Word Cloud Analysis, Frequency Analysis using Pareto Chart, and Topical Models. Word Cloud Analysis provides an immediate visual representation of the most common words, highlighting primary themes. Frequency Analysis uses a Pareto Chart to quantify term frequencies, identifying and graphically representing the most frequently occurring terms to understand their prominence.

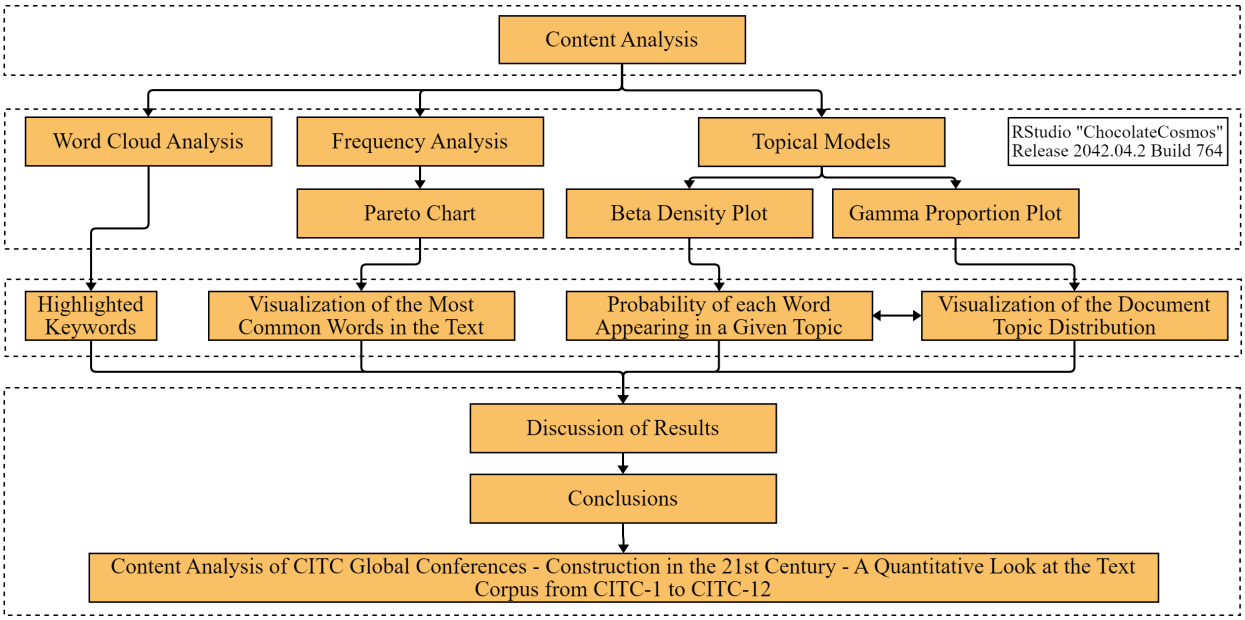


Fig. 1. Methodology Flowchart

Topical Models further analyze the text by identifying key topics and their relevance. This involves the Beta Density Plot, which shows the probability of each word appearing in a given topic, and the Gamma Proportion Plot,

which visualizes the distribution of topics across documents. In this context, a "topic" refers to a cluster of terms that frequently co-occur, indicating a particular theme or subject area discussed in the conference papers. A "document" refers to an individual paper abstract from the CITC conferences. These models help in understanding the spread and focus of topics within the conference materials. The methodology concludes with a Discussion of Results, integrating insights from all analyses, and a summary of key findings, highlighting important trends and themes within the construction industry as presented in the CITC conferences. This structured approach ensures a thorough understanding of the primary themes and priorities discussed in the conferences.

3. Results

3. 1 Word Cloud Analysis

As illustrated in Fig. 2, "Construction" is the most prominent word, indicating that the primary focus of the CITC Global conferences is on the construction industry. Other large words include "project," "management," "research," "building," "cost," "performance," "system," and "study." These words suggest key themes and areas of focus within the construction context. There is a notable emphasis on research and study within the field, reflecting a systematic approach to addressing construction challenges. Additionally, the prominence of terms like housing, infrastructure, sustainability, and environmental impact suggests a strong focus on sustainable development and environmental considerations. Overall, the word cloud illustrates the industry's current priorities, including efficient project management, cost control, and the integration of sustainable practices and technologies.

Fig. 2. Text corpus word cloud (CITC-1 to CITC-12)

3.2 Pareto Analysis

The Pareto chart (Fig. 3) reveals insightful details about the frequency distribution of key terms in the dataset. The x-axis represents the terms, each assigned a serial number, while the y-axis on the left displays the term frequencies. The chart uses bars to indicate the frequency of each term, with taller bars representing higher occurrences. Additionally, the y-axis on the right shows the cumulative percentage, depicted by a red line with points. This line helps visualize the cumulative impact of the terms' frequency, highlighting the most influential ones.

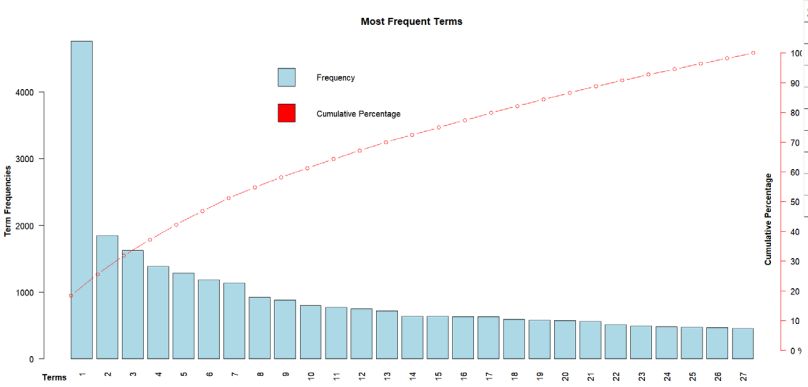
Similar to the word cloud, the most prominent term is "construction" (Serial# 1), which stands out with over 4000 occurrences, significantly higher than any other term. Following this, "project" (Serial# 2) and "management" (Serial# 3) are also highly frequent, each with around 1500 occurrences. Other notable terms include "industry," "study," "projects," "paper," "research," "building," "design," "performance," "cost," "information," "analysis," "system," "data," "factors," "model," "quality," "development," "safety," "risk," "BIM," "results," "based," and "time." These terms, while less frequent than the top three, still play significant roles in the dataset. These terms are divided into the categories of each CITC conference sequence in Table 1. An interesting observation is that BIM had no presence in

CITC-1 but saw a sharp increase by CITC-3 and CITC-12, indicating a significant rise in CITC Global research papers on BIM in 2004 and recent years.

The emergence of terms like BIM highlights evolving research interests over time. The noticeable peaks in BIM-related discussions during CITC-3 and CITC-12 suggest shifts in focus, potentially driven by advancements in technology and its adoption within the industry. To further illustrate the evolution of research interests, emerging terms that have gained attention in recent years should also be considered. In recent years, the concept of digital twins—creating digital replicas of physical assets—has gained significant prominence (Awasthi et al., 2024). BIM is essential for developing digital twins and smart cities, reinforcing its importance in both industry and academia, and preparing us for future technological advancements.

A clear pattern emerges, reflecting the Pareto principle (80/20 rule), where a small number of terms contribute to a large percentage of the occurrences. The cumulative percentage line indicates that the top five terms ("construction," "project," "management," "industry," and "study") account for over 60% of the total frequency. Extending this to the top ten terms, we see that they cover almost 80% of the frequency, highlighting their dominance in the dataset. This concentration of terms suggests that the discourse is heavily centered around a few key concepts, which likely represent the core themes and focal points of the field. This pattern of term frequency distribution highlights the importance of these specific terms in shaping the narrative and indicates that a relatively small set of terms can provide substantial insight into the primary topics and trends within the dataset.

In summary, the Pareto chart illustrates that a small number of key terms, particularly "construction," "project," and "management," dominate the dataset, accounting for a significant majority of the frequency. This distribution suggests that these terms are critical focal points in the CITC-1 to CITC-12 text corpus, reflecting primary themes and areas of interest in the CITC Global conferences. The cumulative percentage line reaffirms the Pareto principle, showing that addressing the top terms can cover the majority of the content's focus, while the long-tail of less frequent terms provides additional, though less critical, context.



Serial#	Term	Serial#	Term	Serial#	Term
1	construction	10	design	19	model
2	project	11	performance	20	quality
3	management	12	cost	21	development
4	industry	13	information	22	safety
5	study	14	process	23	risk
6	projects	15	analysis	24	bim
7	paper	16	system	25	results
8	research	17	data	26	based
9	building	18	factors	27	time

Fig. 3. Pareto Analysis of text corpus (CITC-1 to CITC-12)

Table 1. Frequency of common words in the text corpus (CITC1 to CITC 12)

Most Frequent Terms	CITC1	CITC2	CITC3	CITC4	CITC5	CITC6	CITC7	CITC8	CITC9	CITC10	CITC11	CITC12
construction	499	272	384	373	278	408	305	519	458	380	482	402
project	153	71	97	107	124	205	150	273	206	101	217	143
management	163	83	128	79	96	195	114	268	175	78	128	117
industry	129	108	130	164	66	83	73	148	134	121	107	126
study	60	76	163	187	61	70	68	143	131	106	107	116
projects	84	51	109	93	70	100	82	144	146	58	132	116
paper	128	55	65	49	97	121	75	182	144	72	77	70
research	62	66	103	91	35	55	59	134	121	60	65	76
building	45	51	74	83	46	65	37	166	85	38	87	102
design	65	38	23	73	40	55	69	155	91	48	98	45

performance	74	40	56	53	24	37	99	119	92	61	41	77
cost	59	39	53	33	44	71	44	101	84	57	81	86
information	127	22	54	44	38	87	53	103	43	42	40	67
process	110	25	23	43	38	67	52	92	60	25	62	42
analysis	35	40	57	59	17	75	32	79	69	43	77	55
system	115	20	37	18	39	80	70	125	41	31	16	42
data	46	49	69	69	17	59	38	93	56	38	41	55
factors	23	18	58	42	40	29	39	81	92	60	45	69
model	102	30	42	16	40	40	32	102	50	25	56	49
quality	74	17	52	29	43	65	43	78	78	23	24	48
development	61	52	59	38	61	36	37	54	51	27	34	52
safety	23	69	42	26	15	16	39	39	69	82	42	54
risk	51	4	26	22	12	56	49	86	61	23	51	53
BIM	0	64	109	81	1	1	2	34	30	35	35	90
results	33	23	55	32	13	45	20	82	64	37	30	40
based	39	19	26	37	21	51	35	87	49	27	38	40
time	33	17	33	24	34	52	18	68	49	24	56	47

3.3 Topical Modeling

The topic modeling visualizations from the CITC-1 to CITC-12 conferences (Fig. 4) reveal significant insights into the recurring themes and focus areas within the construction industry. The left side of the visualization displays term-frequency bar charts for four distinct topics, while the right side shows the document-topic distributions. The documents used in this analysis are the paper abstracts from CITC-1 to CITC-12.

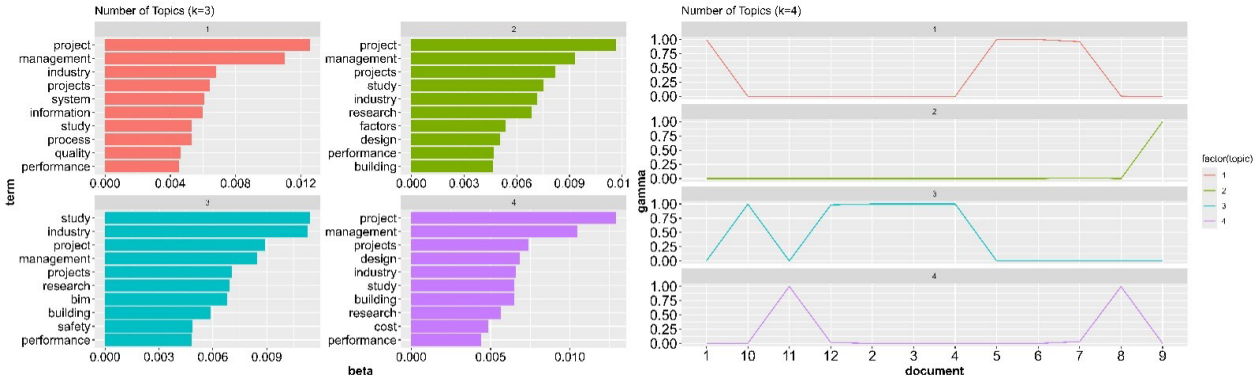


Fig. 4. Topical Modeling of text corpus (CITC-1 to CITC-12)

In *Topic 1*, the top terms include "project," "management," "industry," "projects," "system," "information," "study," "process," "quality," and "performance." The beta values indicate that these terms are strongly associated with this topic. The emphasis here is on project management within the construction industry, highlighting important aspects such as systems, processes, quality, and performance. However, specific technical terms related to project management methodologies (e.g., Agile, Lean Construction), tools (e.g., Primavera, MS Project), and techniques (e.g., Critical Path Method, Earned Value Management) could be incorporated into the corpus to provide deeper insight into the practical methods used in project management.

Topic 2 features terms like "project," "management," "projects," "study," "research," "factors," "design," "performance," and "building." The relatively balanced beta values suggest a comprehensive focus on these areas. This topic emphasizes research and studies within project management, with additional focus on design, performance,

and building factors. The inclusion of specific design methodologies (e.g., Integrated Project Delivery, Design-Build), research tools (e.g., qualitative analysis software, simulation models), and performance metrics (e.g., Key Performance Indicators, Six Sigma) would enhance the understanding of the specific methods and tools that are integral to this topic.

In *Topic 3*, the prominent terms are "study," "industry," "project," "management," "projects," "research," "BIM," "building," "safety," and "performance." The beta values are evenly distributed, indicating a balanced importance across these terms. The focus here is on industry studies, project management, and safety, with a notable inclusion of BIM.

Topic 4 includes terms such as "project," "management," "projects," "industry," "study," "building," "cost," and "performance." The strong beta values for these terms highlight their relevance to this topic. This topic revolves around project management but with an added emphasis on cost and building performance. To provide a more comprehensive view, it would be beneficial to mention cost management techniques (e.g., Life Cycle Cost Analysis, Value Engineering), building performance metrics (e.g., energy efficiency, sustainability ratings), and specific tools (e.g., cost estimation software, Building Performance Simulation tools).

The document-topic distributions on the right side of the visualization show which documents are most strongly associated with each topic. *Topic 1* is strongly represented in documents 1, 10, and 12, indicating a heavy focus on project management and industry-related topics in these documents. *Topic 2* is most associated with document 9, suggesting a primary concern with research and studies within project management. *Topic 3* has high gamma values for documents 1, 4, and 8, pointing to a strong emphasis on industry studies, project management, and safety, particularly with a focus on BIM. *Topic 4* is prominently featured in document 7, which focuses on project management, industry studies, cost, and building performance.

Overall, the topic modeling visualizations provide a detailed overview of the main themes across the CITC Global conferences. Key areas of focus include project management, industry studies, system processes, quality, performance, and safety, with significant attention to research, BIM, and cost considerations. The document-topic distributions further illustrate which documents are most closely associated with each topic, allowing for a targeted analysis of specific conference papers or discussions. This comprehensive overview helps in understanding the recurring themes and priorities within the construction industry as presented at the CITC Global conferences. This detailed thematic analysis helps in identifying key focus areas such as:

- Project management practices
- Research and advancements in design and performance
- The integration of BIM and safety considerations
- Cost management and building performance optimization

These insights are crucial for industry stakeholders, researchers, and policymakers to understand the trends, challenges, and innovations within the construction sector. The ability to pinpoint which documents are most relevant to each topic also allows for targeted analysis, aiding in more efficient literature reviews and knowledge dissemination.

4. Discussion

The comprehensive content analysis of the CITC Global conferences from CITC-1 to CITC-12 reveals significant insights into the evolving landscape of the construction industry. The word cloud analysis highlights "construction," "project," "management," "research," "building," "cost," "performance," "system," and "study" as the most prominent terms. This indicates a strong focus on project management, cost control, and the integration of sustainable practices and technologies within the construction industry. The emphasis on research and study reflects a systematic approach to addressing construction challenges, while terms like housing, infrastructure, sustainability, and environmental impact underline the industry's commitment to sustainable development and environmental considerations.

The Pareto chart analysis further underscores the dominance of terms like "construction," "project," and "management," which together account for a significant majority of the term frequency. This distribution aligns with the Pareto principle, suggesting that a small number of key terms are critical focal points in the conference discussions. These terms, including "industry," "study," "projects," "paper," "research," "building," "design," "performance," "cost," "information," "analysis," "system," "data," "factors," "model," "quality," "development," "safety," "risk," "BIM," "results," "based," and "time," highlight the primary themes and areas of interest within the CITC conferences. The cumulative percentage line reaffirms that addressing these top terms can cover the majority of the content's focus, providing a comprehensive overview of the industry's priorities.

The topic modeling results show that certain themes, such as project management and industry studies, have consistently been at the forefront of discussions. However, the analysis also reveals the under-representation of technical-specific terms representing methods and tools, aside from BIM. Emerging technologies such as digital twins, which have gained traction in recent years, should be included in future analyses to provide a more holistic view of the industry's evolving interests.

5. Conclusions

The content analysis of the CITC Global conferences from 2002 to 2022 (CITC-1 to CITC-12) reveals recurring themes and evolving focus areas within the construction industry. Key areas of emphasis include project management, cost control, sustainable development, and the integration of advanced technologies such as BIM. The word cloud and Pareto chart analyses highlight the industry's commitment to addressing these critical issues through systematic research and study. The topic modeling visualizations provide further insights into the distribution and relevance of various topics, illustrating how different documents are associated with specific themes.

Overall, this analysis offers valuable insights for researchers, practitioners, and policymakers navigating the ever-changing landscape of the construction industry. By understanding the primary themes and priorities discussed in the CITC conferences, stakeholders can better address current challenges and anticipate future trends, ultimately contributing to more effective and informed decision-making processes within the industry. The structured approach to content analysis employed in this study ensures a thorough understanding of the primary themes and provides a foundation for future research and strategic planning in the construction sector.

References

Awasthi, K., Padwekar, K., Misra, S. C., Awasthi, K., Padwekar, K., & Misra, S. C. (2024). Digital Twin: A Unified Definition, Issues, Challenges, and Opportunities (digital-twin) https://doi.org/10.4018/978-1-6684-7366-5.ch074

Bilge, E. Ç., & Yaman, H. (2021). Research trends analysis using text mining in construction management: 2000–2020. *Engineering, Construction and Architectural Management*, *29*(8), 3210–3233. https://doi.org/10.1108/ECAM-02-2021-0107

Cidell, J. (2010). Content clouds as exploratory qualitative data analysis. *Area*, *42*(4), 514–523. https://doi.org/10.1111/j.1475-4762.2010.00952.x

Dicle, M. F., & Dicle, B. (2018). *Content Analysis: Frequency Distribution of Words* (SSRN Scholarly Paper 2997101). https://doi.org/10.2139/ssrn.2997101

Jagannathan, M., Roy, D., & Delhi, V. S. K. (2022). Application of NLP-based topic modeling to analyse unstructured text data in annual reports of construction contracting companies. *CSI Transactions on ICT*, *10*(2), 97–106. https://doi.org/10.1007/s40012-022-00355-w

Li, W., Tang, L. M., Montayre, J., Harris, C. B., West, S., & Antoniou, M. (2024). Investigating Health and Well-Being Challenges Faced by an Aging Workforce in the Construction and Nursing Industries: Computational Linguistic Analysis of Twitter Data. *Journal of Medical Internet Research*. https://doi.org/10.2196/49450

Liu, G., Nzige, J. H., & Li, K. (2019). Trending topics and themes in offsite construction (OSC) research: The application of topic modelling. *Construction Innovation*, *19*(3), 343–366. https://doi.org/10.1108/CI-03-2018-0013

Mohamed, M., & Tran, D. Q. (2022). Content analysis of e-inspection implementation for highway infrastructure construction projects. *Engineering, Construction and Architectural Management*, *30*(7), 2621–2644. https://doi.org/10.1108/ECAM-08-2021-0699

Rashidi, A., Tamošaitienė, J., Ravanshadnia, M., & Sarvari, H. (2023). A Scientometric Analysis of Construction Bidding Research Activities. *Buildings*, *13*(1), Article 1. https://doi.org/10.3390/buildings13010220

Zhang, Q., Hong, Z., & Su, X. (2021). Content Analysis Based on Knowledge Graph: A Practice on Chinese Construction Contracts. In G. Ye, H. Yuan, & J. Zuo (Eds.), *Proceedings of the 24th International Symposium on Advancement of Construction Management and Real Estate* (pp. 823–837). Springer. https://doi.org/10.1007/978-981-15-8892-1_59