

Research Trend Analysis for Construction Automation

Kiyouon Pyo^a, Dongmin Lee^b, Dongyoun Lee^c, Hyunsu Lim^d, Hunhee Cho^e, Kyung-In Kang^f

^{a, b, c, d, e, f}School of Civil, Environmental, and Architectural Engineering, Korea University, Republic of Korea
E-mail: gpyyo@korea.ac.kr, ldm1230@korea.ac.kr, dy_lee@korea.ac.kr, iroze00@korea.ac.kr, hhcho@korea.ac.kr,
kikang@korea.ac.kr

Abstract –

Automation has been an active research topic in the construction industry for many years to improve the accuracy, reliability, and efficiency of the construction process. This study identifies global research trends in automation in construction during the past 15 years by analyzing papers published in the journal *Automation in Construction*. Text mining, data clustering, and keyword network analysis were applied to identify emerging technologies.

According to the clustering of the most recent three years' research, the main topics fall into seven categories: BIM (building information modeling), optimization, data acquisition, big data and AI, 3-D printing, wearable devices, and others. Over the various regions and periods, the results showed that the research focus has shifted from traditional mechanical automation to construction data acquisition, management, and analysis, introducing a new historical period in construction automation.

Keywords –

K-means clustering, Text mining, Keyword network, Trend analysis, Construction automation

1 Introduction

Construction automation can improve construction quality, work productivity, and safety. Under the trend Industry 4.0, studies of smart construction technologies related to machine learning and the internet of things (IoT) and information and communications technologies (ICT) are currently the focus of active research.

Research trend analysis is essential to securing core source technology through R&D, and it is also essential to assess current technology status and to identify promising research areas for construction automation. Therefore, it is important to assess the current state of technology and to identify promising research fields in construction automation [1–5].

As the primary outcome of basic research, a research paper contains detailed and specific

descriptions of the research activities, so papers are rich and valuable data for understanding innovations and changes in technology [6]. Therefore, the use of research data for understanding the status of the technologies of construction automation can be considered as a significant analysis method.

Text mining is a representative, quantitative research method using natural language processing (NLP), and is currently used to extract keywords from large amounts of unstructured data and to derive implications through clustering [7]. In this paper, we apply machine learning based on the K-means clustering method to cluster research papers into several groups covering individual research topics. Core technologies by categorical groups are identified through the retrospective analysis method.

Keyword network analysis can be used in various studies, such as coincidence analysis of main keywords, social network analysis, and papers and patents citation analysis. This technique is currently used in various fields, such as social science, physics, and economics [8]. We used keyword network analysis to follow historical trends and to anticipate future patterns of research keywords.

The purpose of this study is to contribute to proactive research and development by comparing and analyzing research papers related to domestic and foreign construction automation technology in depth. Also, keyword-based semantic approach suggested a possibility of effectiveness in the analysis of technological trends, prospects in the construction field.

2 Theoretical Background

2.1 Literature Review

Several studies on trend analysis and prediction using research papers and patents have been conducted for construction automation. Son and Kim identified global trends and issues in automation and robotics technology in the construction industry by analyzing papers published in the proceedings of the International Symposium on Automation and Robotics in Construction (ISARC) [9]. Balaguer and Abderrahim analyzed

robotics and automation in construction [10].

However, since 2010, there has been no analysis of trends in construction automation research. This study identifies the state-of-the-art research trends in construction automation to reveal promising technologies in the past 15 years.

2.2 Text Mining

Text mining refers to extracting information from unstructured text data in natural languages. It extracts and processes hidden patterns or relationships and finds useful and meaningful information. It is based on NLP. The process of text mining consists of data collection, preprocessing, analysis, and visualization of extracted data.

2.3 TF-IDF

TF-IDF is a weight value for each word used in text mining. It is a statistical weight used to determine the importance of a word in a specific document.

Term frequency (TF) is a value indicating how often the selected word appears in a document. Higher values indicate that the term is more important in the document. If the word is frequently used within a set of documents, it generates a high document frequency. The reciprocal of this value is called the inverse document frequency (IDF). TF-IDF is the product of TF and IDF [11].

The IDF value is determined by the nature of the document family. For example, the word “construction” does not appear very often among general documents, so its IDF value is high and it can be a keyword in documents. However, if the document family is a collection of construction papers, “construction” becomes a cliché, and other words will receive high weight values.

2.3.1 Mathematical Background

The simplest method to calculate the TF is to use the total frequency of the word appearing in the document. If the total frequency of word t in document d is $f(t, d)$, then the simplest TF calculation method is expressed as $TF(t, d) = f(t, d)$.

Another way to calculate the TF value is called increasing frequency, adjusting the word frequency according to the length of the document as follows:

$$tf(t, d) = 0.5 + 0.5 \cdot \frac{f_{t,d}}{\max\{f_{t',d}: t' \in d\}}$$

IDF indicates how frequently a word appears in an

entire document set. It can be obtained by dividing the total number of documents by the number of documents containing the word and then taking the log.

$$idf(t, D) = \log \frac{N}{|\{d \in D: t \in d\}|}$$

If a word appears frequently in a particular document but only a small number of documents contain the word, the TF-IDF value is high. Therefore, the TF-IDF value is utilized to filter words that are common in all documents. Because the IDF log value is always one or more, the IDF value and the TF-IDF value are always zero or more. As the number of documents containing a certain word increases, the value in the log function becomes closer to one. In this case, the IDF value and the TF-IDF value converge to zero:

$$tfidf(t, d, D) = tf(t, d) \times idf(t, D)$$

2.4 Keyword Network Analysis

To obtain a correlation and historical tracking of keywords, we used keyword network analysis. A keyword network consists of main streams, nodes, and links. A keyword is extracted from the author’s keyword list and indexing keywords provided by *Automation in Construction*. This keyword-based analysis creates several groups of keyword networks with relationships between keywords, and the results afford a sense of the evolution of research trends in construction automation.

2.5 Clustering Method

A clustering method is an analytical technique in which objects are grouped into clusters so that individuals having similar characteristics are grouped together.

In text mining, the extracted keywords are vectorized, and the correlations between documents and words are extracted based on a distance, such as Euclidean distance, cosine similarity, or Manhattan distance between words to perform clustering.

The main purpose of data clustering is to identify the characteristic of each clustered class, and K -means clustering, K -nearest neighbors clustering, and principal component analysis are the main methods used to cluster the data.

This study used K -means clustering after text mining. This method clusters a given data set into K

classes according to the median values, and it is an efficient method for finding clusters from large amounts of data.

3 Research Analysis

To analyze the research trends in the automation of construction, three main analyses were carried out in this paper. First, the number of papers published over the past 14 years was surveyed to observe global trends in the research and to find quantitative contributions. Second, studies over the past 15 years were divided into five-year periods, to derive the correlation for each research keyword over time through keyword network analysis. Finally, research methods and subjects were extracted and clustered over the past three years to investigate the latest research trends.

3.1 Data Collection

In this study, 1880 research papers were collected, published by *Automation in Construction*, during the period between 2004 and 2018. An R program was used to conduct keyword extraction, preprocessing, frequency acquisition, and *K*-means clustering.

3.2 Relative Research Contributions by Country

Figure 1 quantitatively shows research trends in the world's automation in the construction area from 2004 to 2017. The graph shows the number of research papers related to construction automation published in each country over those 14 years. The research scope of this paper covers papers published until February 2018; however, the papers published in 2018 were excluded from this analysis to show annual trends.

From 2010 to 2013, a quantitative expansion in research around the overall region is observed, while numbers remained almost steady in the earlier period. The United States' contribution to the total number of papers has been very dominant overall. However, China showed remarkable growth from 2012 to 2013. This is an indirect indicator that China's research investment toward automation construction is growing explosively. It has been predicted since the late 2000s that China will become a leading country in future automation research along with the increasing R&D in construction technology [12].

3.3 Keyword Network

To visualize and analyze the trend flow of construction automation over the past 15 years,

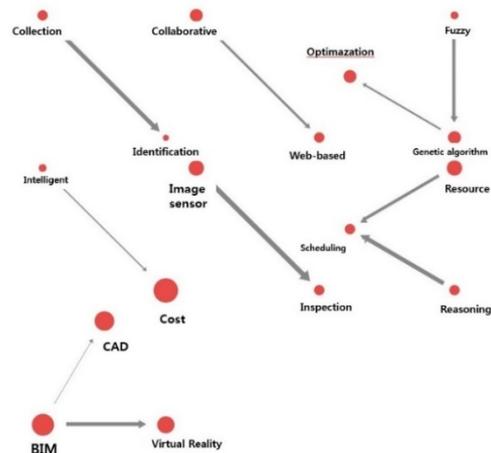


Figure 2. Keyword network, 2004–2008

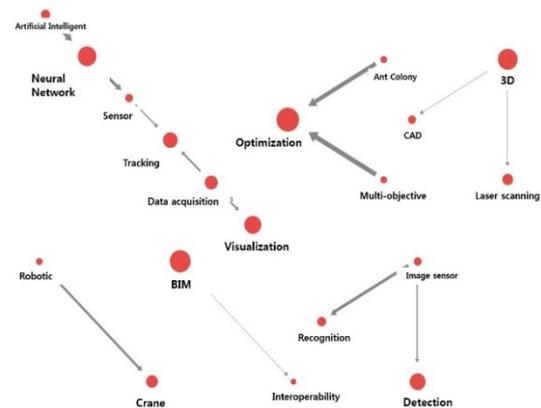


Figure 3. Keyword network, 2009–2013

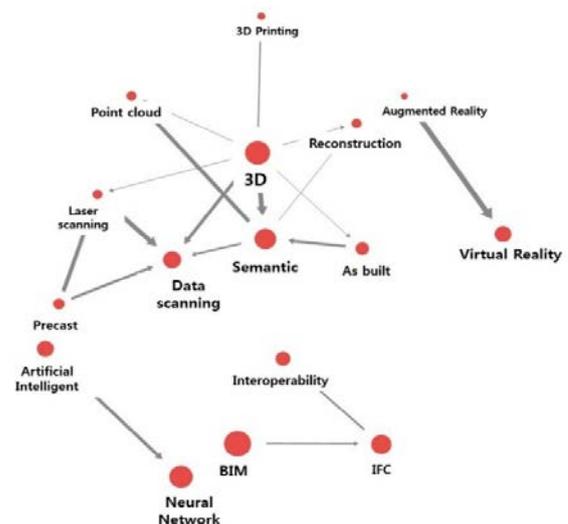


Figure 4. Keyword network, 2014–2018

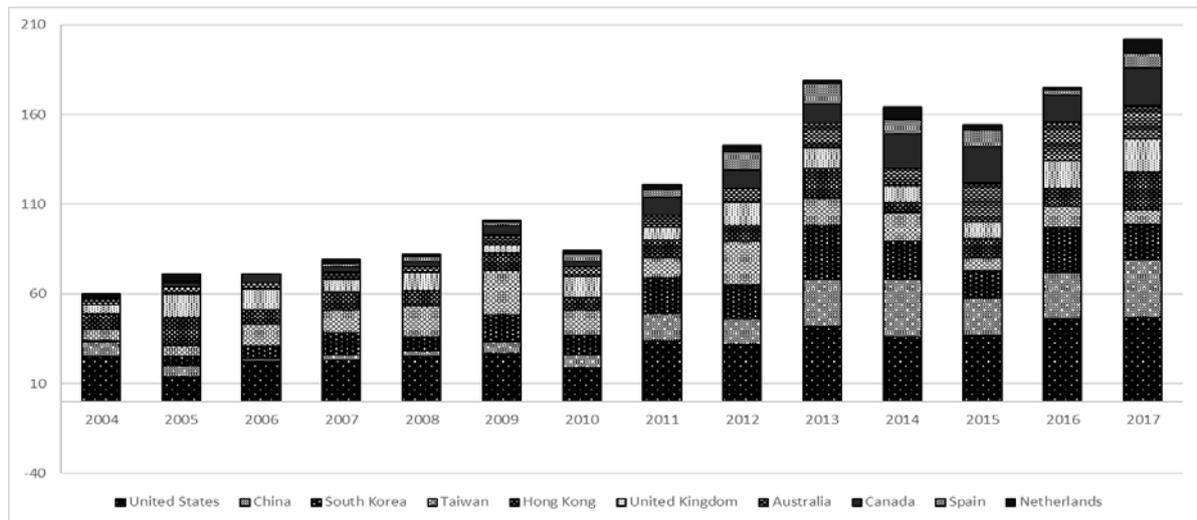


Figure 1. Relative research contributions by country, period 2004–2017

abstracts were gathered in five-year groups, keywords were extracted, and the correlations between keywords are visualized through the following network diagrams. In these diagrams, a larger circle indicates higher frequency and significance of the keyword. The direction of each arrow indicates the dependencies between keywords. A keyword without an arrow link means a separated technical factor keyword.

As shown in Figs. 2–4, research on BIM technology elements is maintained at a high level, and industry foundation classes (IFCs) and interoperability are active research fields. This suggests that the two technologies play important roles both as technology providers and technology adopters in construction automation.

Optimization studies are evolving and proceeding steadily with methodologies such as fuzzy theory, ant colony, genetic algorithm, particle swarm theory, and neural network theory on cost and scheduling, off-site placement, delivery, and simulation.

In the 2014–2018 period, improvements in emerging research areas such as laser scanning and 3-D reconstruction through point-cloud were observed. Further research on the visualization of buildings using drones and conversion between 2-D and 3-D drawing is expected. Real-time detection and tracking of construction equipment and workers with image sensors, augmented reality, virtual reality, and artificial intelligence can be identified as independent technology fields, and these fields are expected to be actively related with construction automation in future.

3.4 Research Trend during the Past Three Years

For research papers in 2015–2017 plus the 113 research papers in early 2018, *K*-means clustering was conducted to extract and analyze the main research topics and keywords for a total of 469 research papers.

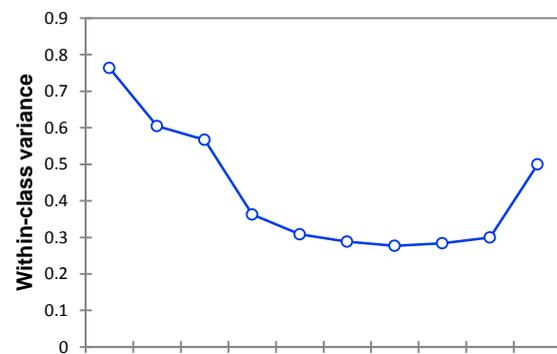


Figure 5. Variance within classes according to the *K*-value in *K*-means clustering

The *K*-value that minimizes the variance of the mean distance for each cluster was selected through repetition. Figure 5 shows the variation of variance according to *K*-value.

The topics and keywords extracted from each group after selecting a *K*-value of 7 are shown in Table 1. The number below the topic title is the number of research papers containing that topic.

Table 1 Extracted topics and keywords of the 2015–2018 group

Topic	Keywords
BIM (110)	CAD, 3D, 4D, management, IFC, data, interoperability, simulation, assessment
Optimization (57)	Algorithm, network, scheduling, layout, multi object, planning, swarm, harmony, genetic, ant

	colony, path
Data acquisition, (56)	Reconstruction, point cloud, detection, image sensing, RTLS, RFID, laser scanning, identification, inspection, monitoring, visualization
Big data, AI (29)	Neural, network, machine learning, energy, IoT, decision, monitoring, surveillance
Wearable devices (13)	Safety, worker, health, device, monitoring, occupational, physiology, sensing, management, hazard, identification
3-D Printing (12)	Fabrication, contour crafting, additive, adhesion, concrete, interlayer, pavement, structural
Other (192)	Energy, decision-making, safety, environment, equipment, earthwork, tunnel, bridge, assembly, mobile, musculoskeletal, prefabricate, excavator, interaction

In the clustering process, six methodology keywords for construction automation were extracted to improve clustering efficiency, and then the scope and objects of the classified categories were analyzed. The topics were divided into seven main categories; BIM, optimization, data acquisition, big data and AI, 3-D printing, wearable devices, and other.

BIM has become a common keyword in most fields, from architecture to civil engineering, and has been actively studied in all fields. Interoperability has been very actively studied to solve the problems of data sharing and consistency through an IFC.

In data acquisition, construction information was acquired through 2-D and 3-D image sensing and laser scanning. Extensive research was conducted on equipment and worker movements through site

monitoring, maintenance of buildings through detection of damage sites, and 3-D reconstruction through point clouds.

Optimization has been applied to various fields, such as off-site yard and form placement, crane and lift operation, critical path analysis, and construction material management using different algorithms. Many studies have utilized experimental and evolutionary methods such as artificial intelligence, neural network theory, and big data.

In addition to the BIM and data acquisition topics, because the several technical elements have converged, it was difficult to specify papers as applying only to one research field. For example, laser scanning and BIM technology have been utilized together to analyze construction errors between as-built and as-designed through comparisons [13].

For wearable devices, studies on safety and risk management through real-time tracking were mainly performed by attaching sensors to workers.

There were also 192 papers not belonging to the above topic categories (grouped as “other”), and their keywords throughout the category were mainly related to construction data.

4 Conclusion

Text mining and *K*-means clustering were applied to research papers to analyze research trends and promising technologies for construction automation. The abstracts of 1880 papers published between 2003 and 2018 were analyzed in this study.

The construction industry mainly depends on manpower and is very vulnerable to variations in conditions. Different regional environments mean that no two projects are identical, which increases the dependency on experience and past performance. This leads to great difficulty in accumulating and utilizing well-organized quantitative construction information data.

Our research has discovered the changes in direction and focus of construction automation, from robotics development and machine automation to IT technologies. Construction robot development was very strong during the 1990s and was expected to be still the main research topic in construction automation [10]. However, only 13 robotics-related papers were found in the clustering analysis of the past three years of research. This was a relatively low research status compared with other topics. This implies that construction automation is not only replacing the labor force by adopting new construction methods but also by acquiring, analyzing, processing, and managing

construction knowledge data using emerging technologies, such as big data processing, IoT, and artificial intelligence technologies.

A limitation of this research is that this paper gathered only abstract data rather than complete research papers; however, it is important that we used quantitative analysis methods rather than a qualitative method such as the Delphi technique or a survey. It was difficult to verify the characteristics and details of each research case. In addition, unique studies were treated as outliers and assigned to clusters that have less correlation in the clustering process.

By analyzing correlations with a keyword network diagram, current research trends and issues in construction automation were identified. In subsequent research, we expect that by tracking the knowledge transfer process through each paper's influence factor and identifying collaborative research networks based on research institutes and countries, quantitative and qualitative analysis of research flow and development will be possible.

Acknowledgements

This research was supported by a grant (18AUDP-B106327-04) from the Architecture & Urban Development Research Program funded by the Ministry of Land, Infrastructure and Transport of the Korean Government.

This work was also supported by the National Research Foundation of Korea through a grant funded by the Korean Government (2016R1A2B3015348). The contribution of the Ministry of Science, ICT and Future Planning is gratefully acknowledged.

References:

- [1] Park Y.-N., Lee, Y.-S. Kim, J.-J., and Lee, T. S. The structure and knowledge flow of building information modeling based on patent citation network analysis. *Automation in Construction*, 87:215–224, 2018.
- [2] N. Li, R. Han, and X. Lu. Bibliometric analysis of research trends on solid waste reuse and recycling during 1992–2016. *Resources, Conservation and Recycling*, 130:109–117, 2018.
- [3] X. Yu and B. Zhang. Obtaining advantages from technology revolution: A patent roadmap for competition analysis and strategy planning. *Technological Forecasting and Social Change*, in press, 2017.
- [4] K. Song, K. Kim, and S. Lee. Identifying promising technologies using patents: A retrospective feature analysis and a prospective needs analysis on outlier patents. *Technological Forecasting and Social Change*, 128:118–132, 2017.
- [5] H. P. Cho, H. Lim, D. Lee, H. Cho, and K.-I. Kang. Patent analysis for forecasting promising technology in high-rise building construction. *Technological Forecasting and Social Change*, 128:144–153, 2018.
- [6] H. Chen, W. Jiang, Y. Yang, Y. Yang, and X. Man. State of the art on food waste research: a bibliometrics study from 1997 to 2014. *Journal of Cleaner Production*, 140:840–846, 2017.
- [7] Y. M. Goh and C. U. Ubeynarayana. Construction accident narrative classification: An evaluation of text mining techniques. *Accident Analysis & Prevention*, 108:122–130, 2017.
- [8] H. Y. Yang. Study on the linkage between the fields of science and technology and characteristics of R&D program using the network analysis. *Korea Institute of S&T Evaluation and Planning*, 2010.
- [9] H. J. Son and C. W. Kim. Trend analysis of research and development on automation and robotics technology in the construction industry. *KSCE Journal of Civil Engineering*, 14(2):131–139, 2010.
- [10] C. Balaguer and M. Abderrahim. Trends in robotics and automation in construction. *Robotics and Automation in Construction*, 2008.
- [11] S. J. Lee and H. J. Kim. Keyword extraction from news corpus using modified TF-IDF. *Journal of Society for e-Business Studies*, 14(4):59–73, 2010.
- [12] Zhang, Xiaoling, Skitmore, Martin, Wu, Yuzhe, Ye, Kunhui. A regional construction R&D evaluation system for China. *Construction Management and Economics*, 28(12):1287–1300, 2010
- [13] V.S. Kalasapudi, P. Tang, and Y. Turkan. Computationally efficient change analysis of piece-wise cylindrical building elements for proactive project control. *Automation in Construction*, 81:300–212, 2017.