The 10th International Conference on Construction Engineering and Project Management Jul. 29-Aug.1, 2024, Sapporo

Advancing Defect Resolution in Construction: Integrating Text Mining and Semantic Analysis for Deeper Customer Experiences

Wonwoo Shin¹, SangHyeok Han² and Sungkon Moon^{3*}

¹ Department of Civil Systems Engineering, College of Engineering, Ajou University, 16499 Suwon, Republic of Korea, E-mail address: <u>wonwdoublo@ajou.ac.kr</u>

² Department of Building, Civil and Environmental Engineering, Concordia University, H3G 1M8, Montréal, Québec, Canada, E-mail address: <u>sanghyeok.han@concordia.ca</u>

³ Department of Civil Systems Engineering, College of Engineering, Ajou University, 16499 Suwon, Republic of Korea, E-mail address: <u>skmoon@ajou.ac.kr</u>

Abstract: According to the South Korean Ministry of Land, Infrastructure, and Transport, instances of defect dispute resolutions, primarily between construction contractors and apartment occupants, have been occurring at an annual average of over 4,000 cases since 2014 to the present day. To address the persistent issue of disputes between contractors and occupants regarding construction defects, it is crucial to use customer sentiment analysis to improve customer rights and guide construction companies in their efforts. This study presents a methodology for effectively managing customer complaints and enhancing feedback analysis in the context of defect repair services. The study begins with collecting and preprocessing customer feedback data. Semantic network analysis is used to understand the causes of discomfort in customer feedback, revealing insights into the emotional sentiments expressed by customers and identifying causal relationships between emotions and themes. This research combines text mining, and semantic network analysis to analyze customer feedback for decision-making. By doing so, defect repair service providers can improve service quality, address customer concerns promptly, and understand the factors behind emotional responses in customer feedback. Through data-driven decision-making, these providers can enhance customer rights and identify areas for construction companies to improve service quality.

Keywords: defect, customer feedback, text mining, semantic network analysis

1. INTRODUCTION

As of 2022, the number of apartments in South Korea totals 12.27 million, accounting for 64% of all housing types [1]. This represents an 18% increase over the past five years [2]. Along with the growth in apartment construction, consumers' perception has shifted from viewing these units merely as living spaces to considering them as commodities [3]. Additionally, there is an increasing interest in the efficient use and management of apartments [4].

According to the Korean Act on the Management of Multi-Unit Residential Buildings, defects are defined as 'flaws due to construction errors that lead to cracks, settlements, breakages, bulging, leaks, etc., which compromise the safety, functionality, or aesthetic integrity of the building or facility.' Furthermore, it is stipulated that 'the constructor is responsible for defect repairs for a period of 2 to 10 years from the date the apartment is handed over to the residents' [5].

However, the lack of objective standards for defect judgment has led to serious social issues, with defects arising during construction, sale, and leasing processes. Since 2014, there have been more than 4,000 disputes annually [6]. The Ministry of Land reports that disputes often escalate into lawsuits due to differences in perception between customers and constructors, leading to socio-economic losses [7]. In fact, the top four construction companies in Korea spent over 300 billion won on defect repair provisions in 2022 [8].

Previous studies have attempted to analyze residents' discomfort through surveys[9, 10]. Specifically, Seshadhri and Paul [9] conducted surveys on the performance and requirements of facilities in university research buildings. However, these approaches had limitations in mining data to identify the causes of users' discomfort.

Therefore, this study aims to analyze the reasons for customers' discomfort based on their feedback, utilizing text mining and semantic network analysis. The study was conducted in the following stages: (1) Mining customer opinions to construct a synonym dictionary. (2) Preprocessing the data based on the constructed dictionary. (3) Extracting preprocessed data and deriving causal relationships of discomfort felt by customers, considering the connectivity of keywords. The opinions were collected through surveys conducted by construction companies after defect repair services. The results of this study can identify the causes of discomfort based on the connectivity of keywords, which can assist in improving the service quality of construction companies. As the opinions were collected after receiving the service, they provide valuable insights into when the discomfort occurred, either before or after the service. This article, following the introduction, introduces in the literature review section researches applying text mining approaches and semantic network analysis in the construction industry.

2. LITERATURE REVIEW

Following the introduction, this article delves into a literature review, highlighting research that applies text mining and semantic network analysis in the construction industry.

2.1. Research Considering the Satisfaction of Apartment Residents

A study was conducted to analyze the satisfaction of apartment residents based on their requirements and to assess performance. Seshadhri and Paul [9] evaluated the efficiency of building facility operations and maintenance through surveys. Ha [11] conducted research by selecting six criteria for Post Occupancy Evaluation (POE) assessment: habitability, convenience, comfort, safety, economy, and sociability. This study made it possible to qualitatively supply POE-based residential environments for mixed-use apartment complexes in Seoul. Although studies have been conducted to analyze the requirements and satisfaction levels of residents, directly identifying the factors that cause discomfort or dissatisfaction among customers remains a challenge. Therefore, this study aims to enhance the understanding of customer experiences and satisfaction based on the data collected.

2.2. Review of Text Mining Approach in Construction Field

Text mining is a process of structuring unstructured text to discover meaningful patterns and insights [12]. The goal is to solve problems like text summarization and classification, information extraction, and information retrieval [13]. To achieve these goals, text mining has been used in various fields such as education, healthcare, finance, and social science to analyze large volumes of text data, using techniques like keyword extraction, topic modeling, and network analysis [14-16].

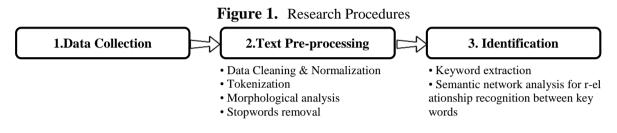
In the construction sector, text mining has been used for analyzing accidents on construction sites and understanding issues in the construction market. For accident analysis, Convolutional Neural Networks (CNN) and Hierarchical Attention Networks (HAN) were trained based on corporate safety reports to understand and prevent accidents [13]. To grasp construction market issues, keyword extraction and visualization were conducted based on text data containing the latest information on the global construction market [17]. Recent research includes a study on Indoor Environmental Quality (IEQ) by conducting text mining of Airbnb accommodation reviews [18]. Term Frequency (TF) and keyword extraction were performed, and based on this, the causes of IEQ-related complaints and trends according to seasonal changes were identified. Therefore, analyzing the connectivity between words through network analysis is necessary.

2.3. Review of the Application of Semantic Network Analysis in Construction Field

Semantic network analysis (SNA) is a technique for creating word networks from unstructured data to understand the relationships between extracted keywords[19]. Keywords are represented as nodes in the network, and links exist between two words when there is a relationship and interdependence between them [20]. In the construction field, semantic network analysis has been used for analyzing causes of delays in construction projects and user opinion analysis in public infrastructure. The analysis of causes for delays in petrochemical projects was based on a report from the Ministry of Industry, Trade, and Resources in Iran [21]. However, this study has the limitation of being confined only to Oil and Gas Producers (OGP). User opinion analysis in public infrastructure utilized complaint data managed by the Ministry of the Interior and Safety in South Korea [22]. Notably, this research introduced a new paradigm focusing on the opinion, who are the most important stakeholders in the lifecycle of urban infrastructure. Therefore, using SNA allows for a better understanding of the relationships between essential keywords.

3. METHODS

Figure 1 shows a summary of the methods in this study. As shown in the figure below, the approach consists of three components. The first step is the collection of user opinions. This was done by gathering feedback from customers at the customer service center of a South Korean construction company. The customers who provided feedback had received repair services for defects that occurred after moving into newly built apartments. Additionally, they assigned scores based on their experience during the service process, which were used to calculate the overall satisfaction. The second step involved extracting information from the collected data and structuring the unstructured data for text mining. This included performing text preprocessing to prepare the data for analysis. Finally, to explore the factors affecting user experience, the preprocessed data was used to calculate TF and conduct SNA. This helped identify the dissatisfaction factors among the residents as keywords and understand the relationships between these factors. Google COLAB was utilized to implement the methods of the proposed research.



3.1. Data collection

In conducting this study, 7,051 customer feedback entries from 23 sites regarding defect repair services conducted between 2018 and 2021 were gathered. The construction company providing the service collected feedback by sending a mobile link to customers. Customers accessing this link provided their opinions about the service in two forms: (1) They rated their satisfaction out of 100 points. (2) They directly wrote additional comments in text form. These opinions were stored in the construction company's central database. For this research, we specifically focused on 831 entries where the satisfaction score was 50 points or lower. Table 1 shows examples of both the text-based customer feedback and the numerical satisfaction scores. The data comprises a total of 13,336 words and 64,673 characters, with an average satisfaction score of 39.67 points. More specifically, customer feedback includes information such as: (1) the type of defect (e.g., wallpaper installation, leakage repair); (2) the customer's emotions regarding the service process and outcome (e.g., "upset", "worst"). In addition to these, although not directly related to the quality of the defect repair, there are comments about the service provided by the construction company's personnel and engineers during the pre- and post-repair process (e.g., response, appointment times).

 Table 1. Examples of the collected raw data

| Raw Data (English translation) | Score |
|---|-------|
| You're so careless about repairing defects | 35 |
| Wait for a long time because you don't keep your appointment. No leak treatment and very poor | 20 |
| response. | 20 |

3.2. Text preprocessing

The purpose of text preprocessing was to transform unstructured collected data into structured data for SNA [23]. Therefore, during the text preprocessing stage, we performed cleaning and normalization, tokenization, and morphological analysis for stopwords removal. Initially, in the data cleaning phase, elements that did not affect the analysis outcome, such as punctuation (e.g., ".", ",", and "?") and indexing numbers, were removed. Subsequently, terms with identical or similar meanings were substituted with their synonyms. As seen in Figure 2, since a hole is a part of apartment defect types, it was replaced with the synonym "defect".

Secondly, the normalized data was tokenized into semantic words, parsing each sentence into individual words. In Figure 2, the sentence "there was a defect in the window so I applied for a defect

repair, but they breach" is tokenized into 17 words: "there", "was", "a", "defect", "in", "the", "window", "so", "I", "applied", "for", "a", "defect", "repair", "but", "they", "breach".

In the third step, morphological analysis was conducted on the tokenized data to identify and extract key terms necessary for the results. For this analysis, we used the OKT module of the KoNLPy Python package, widely used for Korean tokenization [24]. Typically, in Korean, nouns carry essential elements that contain major information about a customer's satisfaction or dissatisfaction with their experience [22]. Also, verbs and adjectives can contain relevant information. Therefore, using the tagging function of the OKT module, nouns, verbs, and adjectives were extracted. As shown in the example in Figure 2 (as shown in red text in Figure 2), four nouns were extracted: "window", "defect", "repair", and "breach". Additionally, two verbs were extracted: "applied", and "delay" (as shown in blue text in Figure 2).

Figure 2. Examples of text pre-processing

• (Korean)2017년2월 창호에 구멍이 있어서, 하자 신청했는데 하자 보수 안 해줌. 얼마나 더 기다려야 하나요? **Raw Data** • (English) In February 2017, there was a hole in the window, so I applied for a defect repair, but they did not repair the defect. How long do I have to wai t? • 2017년2월 창호에 하자이 있어서 하자 신청했는데, 하자 보수 약속불이행. Cleaning 얼마나 더 <mark>지연</mark> 하나요? and • In February 2017, there was a defect in the window, so I applied for a defect Normalization repair, but they breach. How long do I have to delay? • "창호" "에" "하자" "이" "있어서" "하자" "신청" "했는데" "하자" "보수" "약속불이행" "얼마나" "더" "지연" "하나요" • "there" "was" "a" "defect" "in" "the" "window" "so" "I" "applied" "for" Tokenization "a" "defect" "repair" "but" "they" "breach" "How" "long" "do" "I" "have" "to" "delay" • "<u>창호" "에"</u> "하자" <u>"이" "있어서"</u> "하자" "신청" <u>"했는데"</u> "하자" "보수" "약속불이행" "얼마나" "더" "지연" "하나요" Morphological • "there" "was" "a" "defect" "in" "the" "window" "so" "I" "applied" "for" Analysis "a" "defect" "repair" "but" "they" "breach" "How" "long" "do" "I" "have" <u>"to"</u> "delay" • <u>"창호"</u> "에" "하자" "이" "있어서" "하자" "신청" "했는데" "하자" "보수" "약속불이행" <u>"엄마나" "더"</u> "지연" <u>"하나요</u>" **Stopwords Removal** "a" "defect" "repair" "but" "they" "breach" "How" "long" "do" "I" "have" "to" "delay"

Finally, in the stopwords removal phase, very common terms with minimal analytical value were eliminated [25]. The stopwords list is typically created by sorting terms by frequency and manually removing them from the collected data. The stopwords list for the collected data included types of facilities where defects occurred (e.g., "window") and common verbs (e.g., "do", and "was"). In essence, this study focuses not on which defects occurred in the apartment but rather on what factors influenced the dissatisfaction of new apartment residents during the defect repair service process. Therefore, facilities where defects occurred, such as "window" marked in green text in Figure 2, were excluded.

3.3. Identification

3.3.1. Keyword extraction

Keyword extraction is pivotal for identifying the most important words and features in text data, serving as a crucial clue for information extraction, text classification and summarization, and information retrieval [26]. In this study, we extracted keywords representing customer satisfaction and dissatisfaction factors based on TF calculations. TF is a simple and intuitive measure that indicates how frequently a word appears in text data. It is a traditional method of keyword extraction and is widely used due to its straightforward calculation process [22, 23, 27]. To effectively demonstrate these results, we visualized the keywords as a word cloud. This representation intuitively displays the frequency of occurrence within the text data, with keywords having higher TF being displayed in larger font sizes [28].

3.3.2. Relationship recognition between keywords

In this stage, SNA, a key technique in text mining, was implemented to decipher the interconnections among the extracted keywords. SNA's application spans diverse areas, including the analysis of delay causes in construction projects and examining political communications, drawing on unstructured data sources [21, 29, 30]. In the semantic networks we developed, keywords are depicted as "Nodes" and "Edges", graphically representing their interrelationships to simplify comprehension and visualization of the outcomes. Essentially, SNA delves into the nodes and edges forming these networks, facilitating the pinpointing of particular knowledge and information. Text data, being inherently unstructured and composed of words, forms a network of relationships that encapsulate knowledge [19]. Employing SNA in textual analysis, commonly referred to as word network analysis, is instrumental in uncovering these semantic linkages between words and constructing the framework of the text network [31]. In such networks, nodes are equivalent to words, while edges illustrate the inter-word connections.

The foundation of semantic networks typically lies in the co-occurrence relations among words. Co-occurrence is defined as words appearing simultaneously within a sentence, paragraph, or broader text context [32]. In this study, the following phenomenon occurs: for the *i*th word (i = 1, 2, ..., n) and the *j*th word (j = 1, 2, ..., n), $Co(w_iw_j)$ $(i \neq j)$ is calculated as follows.

$$Co(w_i, w_j) = (x+a)^n = \sum_{k=1}^{c} [num(w_i|d_k) \cdot num(w_f|d_k)]$$
(1)

The term $num(w|d_k)$ represents the frequency of word (w) in the kth dataset. Here, d_k signifies the word's occurrence frequency across the entire dataset, and c denotes the total dataset count. Through SNA, we analyzed the co-occurrence frequency of keywords and visualized this network to elucidate the interrelationships among these extracted keywords.

To better comprehend the semantic network's attributes, we calculated its density and centrality. Network density(N) is defined as the ratio of actual node interconnections to the total possible connections, where e(N) symbolizes the count of edges in the network. With N representing the node count, a higher density network facilitates enhanced sharing and information dissemination. Centrality, on the other hand, delineates a node's position relative to the network's central axis. Specifically, the Degree of Centrality (*DC*) mirrors the quantity of nodes that directly connect to a particular node. Thus, a heightened *DC* value signifies extensive connectivity of a node within the network[33]. In this research, both *DC* and the Weighted Degree of Centrality (*WDC*), indicative of maximum centrality, were computed as follows.

$$density(N) = \frac{e(N)}{C(n,2)}$$
(2)

$$DC(node_i) = \frac{e(node_i)}{e(N)}$$
(3)

$$WDC(node_{i}) = \frac{\sum_{j=1}^{n} Co(w_{i}, w_{j})}{\sum_{i=1}^{n} \sum_{j=1}^{n} Co(w_{i}, w_{j})}$$
(4)

4. RESULTS AND DISCUSSION

The outcomes of text preprocessing led to the calculation of TF and the extraction of pivotal keywords. As illustrated in Table 2, the term 'Repair' emerged as the most frequent, followed in sequence by 'Processing', 'Incomplete', 'Promise', and 'Time'. This pattern reveals the customers' discomfort with aspects of the repair service, encompassing its execution, completion standards, promise fulfillment, and time efficiency. Figure 3 showcases a word cloud that graphically represents the TF of the top 43 keywords derived from the dataset.

Table 2. Top 43 words in the 50 points or lower data

| Word | TF value | Word | TF value | Word | TF value |
|----------------|----------|-------------|----------|-----------------|----------|
| Repair | 943 | Company | 64 | Diligence | 27 |
| Processing | 417 | Coordinator | 52 | Unpleasantness | 25 |
| Incomplete | 370 | Work | 49 | Dissatisfaction | 23 |
| Promise | 335 | Swiftness | 47 | Disappointment | 23 |
| Time | 300 | Clean | 47 | Staff | 23 |
| Fulfillment | 293 | Clean-up | 42 | Improvement | 16 |
| Request | 262 | Customer | 41 | Noise | 16 |
| Response | 205 | Replacement | 41 | Inspection | 15 |
| Materials | 148 | Evaluation | 39 | Kindness | 14 |
| Register | 146 | Delay | 39 | Satisfaction | 13 |
| Quality | 126 | Thanks | 37 | Mess | 13 |
| Visit | 99 | Check | 34 | Brand | 11 |
| Contact | 89 | Faulty | 30 | Accuracy | 11 |
| Unsatisfactory | 82 | Finish | 29 | · | |
| Engineer | 79 | Attitude | 27 | | |

Figure 3. Word clouds with top 43 keywords



To map out the interplay among these extracted keywords, we utilized SNA to form a word network grounded on the 43 words exhibiting the highest co-occurrence rates. The constructed word network comprised 20 nodes and 55 edges, with an aggregate of 6,523 co-occurrences and a calculated density of 0.289. Table 3 provides a glimpse into the word interrelations extracted from this network. Remarkably, the pair 'Incomplete + Repair' demonstrated the most frequent co-occurrence (518 instances), signifying their joint presence in 518 data entries.

| Word relationship | Co-Occurrence |
|-----------------------|---------------|
| Incomplete + Repair | 518 |
| Time + Repair | 516 |
| Repair + Promise | 451 |
| Repair + Processing | 438 |
| Repair + Fulfillment | 376 |
| Promise + Fulfillment | 355 |
| Repair + Materials | 347 |
| Repair + Request | 319 |
| Repair + Quality | 306 |
| Time + Promise | 297 |

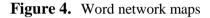
Table 3. Examples of word relationships derived from word networks

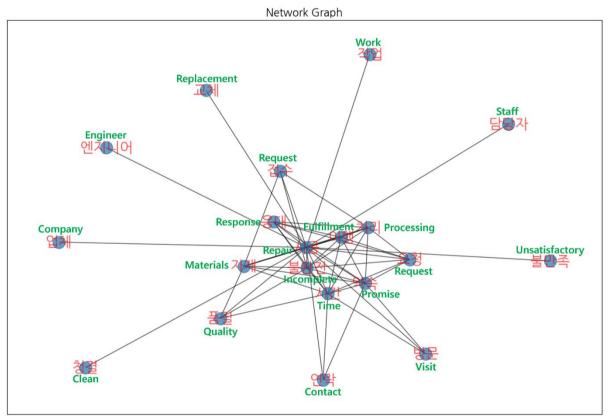
We then proceeded to calculate degree centrality to pinpoint the core words within the word network. As delineated in Table 4, six words demonstrating significant centrality and weighted centrality are cataloged. Notably, both 'Incomplete' and 'Time' shared an identical DC of 0.976, yet 'Incomplete' boasted a higher WDC at 0.195, surpassing 'Time's WDC of 0.163.

| Word | DC | Word | WDC |
|------------|-------|------------|-------|
| Incomplete | 0.976 | Repair | 0.201 |
| Time | 0.976 | Incomplete | 0.195 |
| Repair | 0.952 | Time | 0.163 |
| Promise | 0.929 | Promise | 0.101 |
| Request | 0.905 | Response | 0.095 |
| Response | 0.881 | Request | 0.082 |

Table 4. Top six words with high centrality in the word networks

Figure 4 presents a 2D map visualization of the word network, centering on pivotal keywords like 'Repair' and 'Incomplete'. Examining the words linked to 'Repair', it is evident that residents' experiences are tied to concepts like 'Clean', 'Request', 'Promise', 'Time', and 'Fulfillment'. This implies challenges related to the cleanliness assurance during repairs, the complexity of requesting services, and issues concerning repair timelines and promise realization. Conversely, focusing on another central word, 'Incomplete', we observe connections to 'Quality', 'Materials', 'Request', and 'Process'. This indicates perceptions of compromised material quality in incomplete services and inefficiencies in processes due to incomplete requests.





5. CONCLUSION

In South Korea, disputes over defects between construction contractors and apartment residents are occurring at an annual average of more than 4,000 cases. Apartments, which currently make up 64% of the country's housing types, are seeing a rapidly increasing share. Consequently, it is anticipated that the number of disputes related to defects will rise in the future. In this context, it is crucial to foster efforts to enhance consumer rights and understand the customer experience from the perspective of construction contractors. Thus, this study aims to identify issues arising during the defect repair process. Based on the results of TF analysis conducted in this study, elements that caused discomfort to customers were extracted as keywords and represented in a Wordcloud (e.g., 'Repair', 'Processing', 'Incomplete', 'Promise', and 'Time'). Subsequently, SNA was employed to understand the relationships between these keywords, which were then visualized on a 2D map. The study identified customer requirements for (1) cleanliness, adherence to time commitments, and simplification of service requests

in relation to 'Repair'; and (2) complete service and fairness, underscoring the need for efficient service delivery related to 'Incomplete'.

The methodologies and findings of this study suggest that providers of defect repair services should enhance service quality, address customer issues in a timely manner, and grasp the underlying factors of emotional responses in customer feedback. By adopting a data-driven approach to decision-making, these service providers can bolster consumer rights and pinpoint opportunities for construction companies to ameliorate service quality.

ACKNOWLEDGMENTS

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science and ICT (Grant Number: 2022R1F1A1074039)

REFERENCES

- 1. Korea Real Estate Board. *Current status of apartment*. 2023.
- 2. Statistics Korea. *Population and Housing Census*. 2023; Available from: <u>https://kostat.go.kr/ansk/</u>.
- 3. Son, S.H., et al., *A study on the types and causes of defects in apartment housing*. Journal of the Korea Institute of Building Construction, 2020. **20**(6): p. 515-525.
- 4. Park, J.-M., O.-K. Kim, and J.-L. Kim, *Revitalization of the conciliation system for defect disputes related to apartment buildings-On the technical issue.* Journal of the Korea Institute of Building Construction, 2011. **11**(3): p. 208-220.
- 5. Ministry of Government Legislation. *Multi-Family Housing Management Act*. 2023; Available from: <u>https://www.law.go.kr/</u>.
- 6. Korea Ministry of Land, I.a.T., *Casebook of Dispute Mediation for Defect Review*. 2022.
- 7. Korea Ministry of Land, I.a.T. *Civil complaints forum*. 2023; Available from: <u>https://eminwon.molit.go.kr/main.do?</u>
- 8. Financial Supervisory Service. *Data Analysis, Retrieval and Transfer System(DART)*. 2022; Available from: <u>https://dart.fss.or.kr/</u>.
- 9. Seshadhri, G. and V.K. Paul, *User requirement related performance attributes for government residential buildings*. Journal of Facilities Management, 2017. **15**(4): p. 409-422.
- 10. Abdul-Rahman, H., et al., *Case Study of Facility Performance and User Requirements in the University of Malaya Research and Development Building*. Journal of Performance of Constructed Facilities, 2015. **29**(5): p. 04014131.
- 11. Ha, M.J., A study on the POE (Post Occupancy Evaluation) according to the residential environment of mixed-use apartment complexes in Seoul. International Journal of High-Rise Buildings, 2020. **9**(2): p. 197-212.
- 12. IBM, What is text mining? 2023.
- 13. Baker, H., M.R. Hallowell, and A.J.-P. Tixier, *Automatically learning construction injury precursors from text*. Automation in Construction, 2020. **118**: p. 103145.
- 14. Ferreira-Mello, R., et al., *Text mining in education*. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 2019. **9**(6): p. e1332.
- 15. Lee, J., et al., *BioBERT: a pre-trained biomedical language representation model for biomedical text mining*. Bioinformatics, 2020. **36**(4): p. 1234-1240.
- 16. Macanovic, A., *Text mining for social science–The state and the future of computational text analysis in sociology.* Social Science Research, 2022. **108**: p. 102784.
- 17. Moon, S., et al., *Document management system using text mining for information acquisition of international construction*. KSCE Journal of Civil Engineering, 2018. **22**: p. 4791-4798.
- 18. Villeneuve, H. and W. O'Brien, *Listen to the guests: Text-mining Airbnb reviews to explore indoor environmental quality.* Building and Environment, 2020. **169**: p. 106555.
- Popping, R., *Knowledge graphs and network text analysis*. Social science information, 2003.
 42(1): p. 91-106.
- 20. Motter, A.E., et al., *Topology of the conceptual network of language*. Physical Review E, 2002. **65**(6): p. 065102.

- Zarei, B., H. Sharifi, and Y. Chaghouee, *Delay causes analysis in complex construction projects: A semantic network analysis approach*. Production Planning & Control, 2018. 29(1): p. 29-40.
- 22. Chang, T., S. Chi, and S.-B. Im, *Understanding User Experience and Satisfaction with Urban Infrastructure through Text Mining of Civil Complaint Data.* Journal of Construction Engineering and Management, 2022. **148**(8): p. 04022061.
- 23. Moon, S., et al., Automated Construction Specification Review with Named Entity Recognition Using Natural Language Processing. Journal of Construction Engineering and Management, 2021. **147**(1): p. 04020147.
- 24. Park, L., *KoNLPy: Korean NLP in Python*. 2014.
- 25. Zou, Y., A. Kiviniemi, and S.W. Jones, *Retrieving similar cases for construction project risk management using Natural Language Processing techniques*. Automation in construction, 2017. **80**: p. 66-76.
- 26. Berry, M.W. and J. Kogan, Text mining: applications and theory. 2010: John Wiley & Sons.
- 27. Baek, S., W. Jung, and S.H. Han, A critical review of text-based research in construction: Data source, analysis method, and implications. Automation in Construction, 2021. **132**: p. 103915.
- 28. Sun, J., et al., *Text visualization for construction document information management*. Automation in construction, 2020. **111**: p. 103048.
- 29. Lehmann, F., *Semantic networks*. Computers & Mathematics with Applications, 1992. **23**(2-5): p. 1-50.
- 30. Xiong, Y., M. Cho, and B. Boatwright, *Hashtag activism and message frames among social movement organizations: Semantic network analysis and thematic analysis of Twitter during the# MeToo movement.* Public relations review, 2019. **45**(1): p. 10-23.
- Yoo, M., S. Lee, and T. Ha, Semantic network analysis for understanding user experiences of bipolar and depressive disorders on Reddit. Information Processing & Management, 2019. 56(4): p. 1565-1575.
- 32. Garc ía, M.C.F., et al., *Semantic network analysis of sustainable development goals to quantitatively measure their interactions*. Environmental Development, 2021. **37**: p. 100589.
- 33. Jeon, S.-W. and J.-Y. Kim, *An exploration of the knowledge structure in studies on old people physical activities in Journal of Exercise Rehabilitation: by semantic network analysis.* Journal of exercise rehabilitation, 2020. **16**(1): p. 69.