

Article

Identification of Occupant Dissatisfaction Factors in Newly Constructed Apartments: Text Mining and Semantic Network Analysis

Seok-Ho Noh ¹, Inho Jo ¹ , SangHyeok Han ², Sungkon Moon ^{3,*} and Jae-Jun Kim ¹ 

¹ Department of Architectural Engineering, Hanyang University, Seoul 04763, Republic of Korea; koaroo@naver.com (S.-H.N.); odd.innocent@gmail.com (I.J.); jjkim@hanyang.ac.kr (J.-J.K.)

² Department of Building, Civil and Environmental Engineering, Concordia University, Montréal, QC H3G 1M8, Canada; sanghyeok.han@concordia.ca

³ Department of Civil Systems Engineering, College of Engineering, Ajou University, Suwon 16499, Republic of Korea

* Correspondence: sungkon.moon@gmail.com

Abstract: With apartment buildings representing a rapidly growing share of the residential market in South Korea, the effect of construction defects throughout the life cycle of construction projects, and particularly during the occupancy stage, has emerged as a significant social issue that may ultimately lead to an increase in defect disputes between new occupants and general contractors. An important step toward mitigating the likelihood of these defect disputes is to identify and address the factors that give rise to occupant dissatisfaction during the defect repair process. However, a reliable method by which to identify these factors has yet to be developed. In this respect, the main objective of the research presented in this paper is to develop a method for identifying occupant dissatisfaction factors in the construction defect repair stage. The developed method comprises the following procedures: (i) text pre-processing, which involves data cleaning, normalization, tokenization, morphological analysis, and removal of stopwords; (ii) term frequency-inverse document frequency for keyword extraction; and (iii) semantic network analysis to recognize relationships between words. The method was implemented using a dataset of 12,874 comments in Korean text format obtained from apartment building occupants. Based on the processing and analysis of this dataset, the occupant dissatisfaction factors were found to be: (i) inaccurate and inadequate repair work (represented by such keywords as “Repair”, “Visit”, and “Accuracy”); (ii) failure to keep promises (e.g., “Fulfillment”, “Promise”, and “Change”); and (iii) unprofessional conduct on the part of representatives in the repair service center (e.g., “Response”, “Attitude”, and “Receipt”).

Keywords: occupant dissatisfaction; semantic network analysis; repairing; construction defects; apartment buildings; text mining



Citation: Noh, S.-H.; Jo, I.; Han, S.; Moon, S.; Kim, J.-J. Identification of Occupant Dissatisfaction Factors in Newly Constructed Apartments: Text Mining and Semantic Network Analysis. *Buildings* **2023**, *13*, 2933.

<https://doi.org/10.3390/buildings13122933>

Academic Editor: Md Morshed Alam

Received: 12 October 2023

Revised: 8 November 2023

Accepted: 20 November 2023

Published: 24 November 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The construction industry in South Korea accounts for 14.6% of the country’s gross domestic product (GDP) [1]. In turn, apartments represent 78% of the residential market in South Korea. In this regard, defects occurring during the construction and/or occupancy of apartments have emerged as a serious social issue. According to a report by the Defect Examination Dispute Resolution Committee (DEDRC), a division of South Korea’s Ministry of Land, the number of construction defect disputes in the occupancy stage has been increasing year over year [2]. For example, of the 16,599 total dispute cases filed by the DEDRC from 2017 to 2020, the annual totals were 4089 (2017), 3818 (2018), 4290 (2019), and 4402 (2020). Worldwide, although reliable statistics related to construction defect disputes/litigations are not readily available, recent studies report that the frequency of construction defect litigation is significantly increasing [3–5]. An increase in construction

defect disputes leads not only to wasted time and psychological damage for occupants due to the inconvenience and discomfort caused, but also to increased project costs and damaged reputation for construction companies (i.e., general contractors). In this paper, “defect” refers to something that prevents the facility from functioning after completion in the manner it was intended at the time of executing the contract [4]. In this respect, construction defect disputes may be generated for various reasons, such as incomplete construction work (e.g., tile grouting, fixtures, coats of paint), improper facility design, noncompliance with building codes, low service quality in the repair of construction defects, and poor facility maintenance during the occupancy stage [3,6,7]. This paper focuses on construction defects repaired during the occupancy stage.

Numerous studies have sought to identify the impact of defects on the satisfaction of residential occupants based on surveys, questionnaires, interviews, and statistical methods (e.g., Pearson chi-square method) [8–11]. In this regard, Million et al. [11] found that the inability of the construction company to satisfy occupant requirements is the most significant factor in terms of its adverse effect on occupant satisfaction. Construction defect disputes often arise from occupant dissatisfaction caused by miscommunication between the construction company and the occupant, or by poor-quality construction defect repair (e.g., noncompliance with occupants’ requests). In an effort to improve both service quality and occupant satisfaction, researchers have explored occupant dissatisfaction factors based on occupant complaint data for such facilities and infrastructure as buildings, bridges, water distribution systems, and metro systems [12–15]. Assaf and Srour [16], in a study seeking to develop a building maintenance strategy to enhance occupant satisfaction and building performance, proposed a neural network approach to predict building occupants’ complaints. Despite these efforts to identify the factors influencing occupant satisfaction and dissatisfaction for various types of facilities, previous studies have the following limitations: (i) no systematic study has yet identified factors affecting occupant dissatisfaction in the construction defect repair process for newly constructed apartments, even though occupants of newly constructed apartments who benefit from the defect repair process may have different interests and requirements from those in existing apartments; (ii) there is a lack of resources to provide knowledge of where and how to improve the construction defect repair process for better quality and serviceability of the construction defect repair process due to limitation (i); and (iii) existing methods that use text-mining techniques may not be suitable for efficiently and accurately extracting the desired information related to main sources of occupant dissatisfaction by mining the contents of complaint data in a Korean text format, due to the unique features of the natural language used in complaints.

To address these limitations, the aim of the research presented in this paper is to identify the main factors that result in occupant dissatisfaction for residents moving into newly constructed apartments in terms of the construction defects repair process, based on the occupants’ complaints represented in a Korean text format, so that the results can be used to not only improve quality and serviceability in the construction defect repair process and the level of occupant satisfaction but also mitigate causes leading to construction defect disputes in apartment projects in South Korea. To achieve this objective, the following procedures were undertaken: (i) text pre-processing, which involves cleaning, normalization, tokenization, morphological analysis, and removal of stopwords; (ii) term frequency (TF)–inverse document frequency (IDF) for keyword extraction; and (iii) semantic network analysis (SNA) to recognize relationships between keywords. As a case study, this paper uses occupant complaint data (in Korean text format) obtained from a collaborating construction company in South Korea, this complaint data having been collected directly from occupants during/after the process of construction defect repair.

This paper consists of the following structure: (i) Section 2 discusses the status of studies in occupant satisfaction and dissatisfaction and text mining; (ii) Section 3 describes the methods to identify occupants’ dissatisfaction factors concerning construction defect repairs; (iii) representation and discussion of results analyzed by these methods are given in Section 4; and (iv) the conclusion is presented in Section 5.

2. Literature Review

2.1. Related Work on Occupant Satisfaction and Dissatisfaction

Construction projects can be broadly classified into four types: residential, commercial, industrial, and infrastructure facilities. In the context of South Korea, among these types of construction projects, residential buildings, especially apartment buildings, have received considerable attention due to the increasing incidence of construction-defect-related disputes year over year [2]. In this respect, previous studies investigating construction dispute litigation have reported that most of the common construction defects are related to building envelope issues (e.g., water distribution systems) [17–19]. Construction defect disputes have also been assessed to identify root causes and triggers of construction defects, such as improper project monitoring and control, improper installation methods, poor site-working conditions, and poor design decisions [20,21]. However, Chong and Low [21] have pointed out in this regard that the construction defects occurring in the occupancy phase should be managed differently from ones found in the construction stage since many building defects are latent in nature and do not appear early in the construction stage. In this respect, to identify root causes of the construction defects in the occupancy phase leading to construction defect disputes, some researchers have studied the impacts of construction defects on occupant dissatisfaction [10,11]. These studies have identified incompetence and low-quality service on the part of the construction companies completing the repairs as the main causes of occupant dissatisfaction. Given that the construction companies carrying out the repairs often fall short of satisfying the occupants' requirements and/or expectations, to improve occupant satisfaction it is essential to identify occupant requirements and/or expectations in detail prior to carrying out the repairs. As a preliminary step in this direction, Bazzan et al. [22] proposed an information management model that provides a structured database in which to store text-based resident complaints efficiently and effectively using artificial intelligence techniques. With respect to building maintenance management, a number of studies have taken the approach of examining occupant complaints and satisfaction as the basis for improving building performance strategies (e.g., heating, cooling, and elevators) and indoor environment quality in accordance with the results of interviews, text mining, and neural network analysis [9,16,23–25]. In this respect, Roumi et al. [26] developed a weight system involving indoor environment quality parameters for occupant satisfaction and energy efficiency based on survey results of surveys reflecting the opinions and experiences of the occupants. Abdul-Rahman [27] introduced building performance requirements to improve building facility maintenance and thereby boost user satisfaction. A few studies have undertaken to increase service quality and user satisfaction with public infrastructure through effective infrastructure maintenance [12–14]. Chang et al. [15] recently sought to better understand user satisfaction/dissatisfaction factors by mining the contents of user complaints regarding tunnel and bridge infrastructure.

2.2. Text-Mining Approaches

Text mining identifies meaningful patterns, information, and new insights by transforming unstructured text data into a structured format as defined by the user [28,29]. That is, the purpose of text mining is to extract useful and interesting information and knowledge from a large set of textual data in an efficient and effective manner. Researchers have applied various text-mining techniques, including keyword extraction, word network analysis, topic modeling, sentiment analysis, and opinion mining, to analyze large volumes of textual documents in fields ranging from finance to education and health science [30–32]. In the context of construction management, text mining is used to extract information, classify text, and discover trends or patterns in text from contractual documents, public opinions obtained by survey or questionnaire response data, complaints, accident reports, and other documents (e.g., CAD documents). In this regard, previous studies have clustered and classified project-related documents and site accidents based on keywords and textual similarities [33,34]. Yu and Hsu [35] proposed a content-based text-mining technique to extract the textual content of a CAD document using similarity matching based on application of

vector space modeling. Akanbi and Zhang [36], meanwhile, proposed a novel semantic natural language processing (NLP)-based method that extracts design information (e.g., material type) from construction specifications. This extracted information, in turn, was used in their study to support construction cost estimation.

Given the ubiquity of the Internet, government entities are able to easily solicit public opinion on the condition and level of service of public infrastructure assets through social network services (SNSs) such as posts on blogs, on Instagram, on Facebook, and on Twitter, to name a few. However, it can be difficult to extract the desired information, since datasets from these sources are generally large, unstructured, heterogeneous, and contain data noise. To address this limitation, Zhou et al. [37] proposed an analytical framework that employs topic modeling and sentiment analysis to recognize public opinions on infrastructure megaprojects obtained from social media platforms. In another study, a convolutional neural network was used for deep-learning-based classification of building quality problems [38]. In that study, building quality complaint text data was labeled and the complaints classified automatically based on complaint subjects involving “leakage”, “hollowing”, and “cracking”. In another study, to improve indoor environmental quality (IEQ) in temporary accommodations (i.e., Airbnb), Villeneuve and O’Brien [39] identified seasonal trends in IEQ by quantifying the frequency of multi-domain IEQ complaints based on the contents of Airbnb reviews.

Text mining has also been used for text clustering and text classification within the construction domain. Text clustering assigns each document into one or more groups in a manner that maximizes intra-cluster similarity and minimizes inter-cluster similarity for construction safety analysis and identification, safety accidents, and information retrieval purposes [39–42]. In this respect, Tixier et al. [43] developed a machine-learning model that can predict construction injury types, energy types, and body parts based on analysis of construction injury reports using NLP. To identify construction accidents and their causes, previous studies have proposed knowledge management systems using NLP techniques such as rule-based and conditional random field methods and deep learning techniques [44,45]. Similar studies have classified and identified not only types of job hazards and site accidents but also sources of project risk and human errors based on unsupervised machine-learning techniques (e.g., symbiotic gated recurrent unit (SGRU) and support vector machine (SVM)) [34,46–48]. In a recent study, D’Orazio et al. [49] developed a maintenance severity ranking system that supports decision making regarding prioritization of end-user maintenance requests.

With regard to information extraction, the construction industry has given attention to automation compliance checking (ACC), which supports experts in ensuring regulatory compliance (i.e., adherence of construction work and building design to the relevant codes, laws, contractual obligations, and policies) [50,51]. For example, Zhong et al. [38] introduced an ontology-based semantic modeling approach that investigates regulatory constraints for construction quality inspection and evaluation.

In summary, certain construction defects (e.g., issues with the building envelope performance) may become apparent during the occupancy phase. Consequently, it becomes imperative to address these defects efficiently and accurately to prevent construction-defect-related disputes. Failure to do so can result in occupant dissatisfaction and the emergence of construction disputes. However, most studies have primarily focused on identifying and rectifying the root causes of construction defects appearing in the construction phase, as well as understanding occupant satisfaction and dissatisfaction to enhance existing infrastructure and building maintenance management. A few studies [10,11] have reported that incompetence and low-quality service in the construction defect repair process are significant contributors to occupant dissatisfaction. However, these findings may lack comprehensive information, such as that relating to attitudes of representatives and regarding clean-up, which is essential to develop innovative strategies for improvement of construction defect repair procedures and services. That is, no study has yet identified the factors contributing to occupant dissatisfaction and how and where to enhance the construction

defect repair process. In addition, these factors may differ between occupants moving into newly constructed facilities (such as the apartments in this paper) and those residing in existing ones, as their expectations and requirements regarding the construction defect repair process can vary. Notably, previous research has not distinguished these factors among occupants in new and existing apartments. To address these challenges, this paper aims to identify the factors contributing to occupant dissatisfaction among those living in new apartments when they receive construction defect repair services. This analysis is based on mining of occupant complaint data in text form, offering valuable insights into improving the construction defect repair process.

3. Methods

A summary of the methods employed is provided in Figure 1. The approach is composed of three components: data collection, text pre-processing, and identification of occupant dissatisfaction factors. The first component involved collecting occupants' comments in text format from customer service centers operated by the collaborating construction company in South Korea. The functions of this customer service center are to field complaints about construction defects from occupants of newly built apartments, request sub-contractors to repair them, and schedule the appointments for contractors to diagnose and repair the defects. The customer service team also follows up with occupants who receive repair services to gauge the extent to which they are satisfied with the repair services provided by the contractor, and stores this input in a central database. The second component of the approach involved implementing text pre-processing to transform unstructured data into a structured format for text mining (e.g., SNA) and information extraction purposes. Finally, in the third component of this approach, TF-IDF calculation and SNA were executed to identify and analyze occupant dissatisfaction factors as keywords and characterize their relationships. The research presented in this paper used Python version 3.7.0 in implementing this approach.

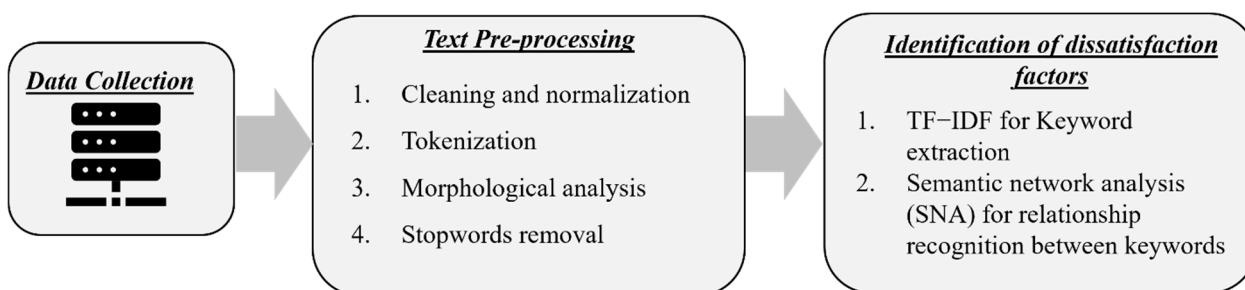


Figure 1. Research methods.

3.1. Data Collection

Once the defects are repaired successfully, the customer service center invites the occupants who have received the repair service to provide feedback through a mobile application, called voice of customer (VC), consisting of two components: (i) evaluating the quality of the repair works based on four levels of satisfaction (i.e., "Very Satisfied", "Satisfied", "Dissatisfied", and "Very Dissatisfied"); and (ii) writing their experiences in a text format representing positive and negative sentiments, depending on the degree of satisfaction/dissatisfaction on the part of the occupants. This VC helps to minimize data bias and prevent manipulation of the textual data by bots since all data are generated by occupants directly. Then, all occupant feedback in the VC is stored in a central database of the custom service center. Consequently, the dataset utilized in this research encompassed a total of 101,387 data points recorded between January 2020 and February 2023. In view of geographical data distribution in South Korea, as depicted in Figure 2, there are a total of five regions: Seoul, Gyeonggi, Kangwon, the Central region, and the Southern region. In terms of population considerations, Seoul stands out with the highest numbers of reported

defects (41,745 data points) and complaints (4609 data points) from 49 sites. Following closely is Gyeonggi, with the second-highest numbers of reported defects (36,806 data points) and complaints (4544 data points). Although the numbers of sites and reported defects are similar in the Central and Southern regions, the Central region registered a higher number of complaints in the context of construction defect repair services. Kangwon, owing to its smaller population compared to other regions, has the fewest data points. Since the primary objective of this research paper is to identify and analyze factors leading to occupant dissatisfaction, thereby enabling improvements in the quality and level of repair services, data points corresponding to occupant ratings of “Very Satisfied” and “Satisfied” have been excluded from the dataset. This exclusion resulted in 7907 data points corresponding to occupant ratings of “Dissatisfied” and 3243 data points corresponding to occupant ratings of “Very Dissatisfied.”

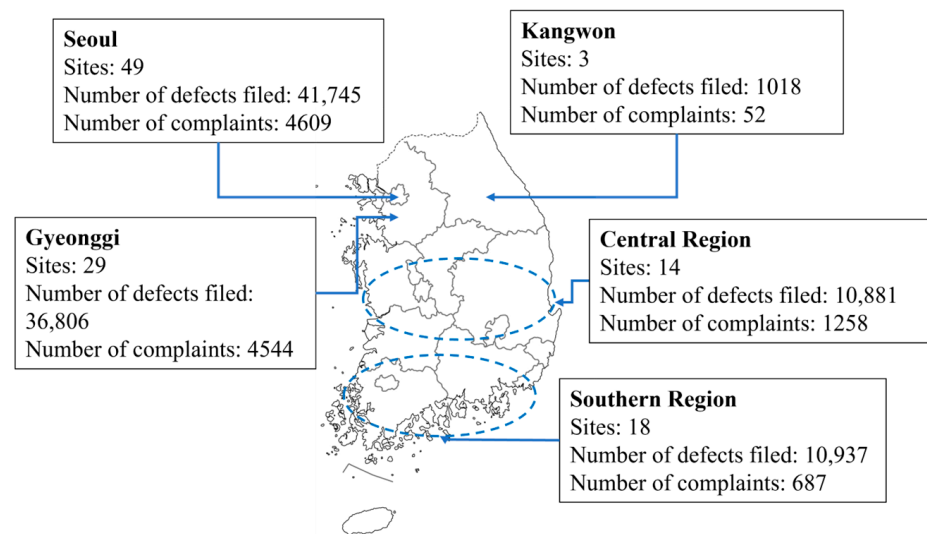


Figure 2. Distribution of collected data in South Korea.

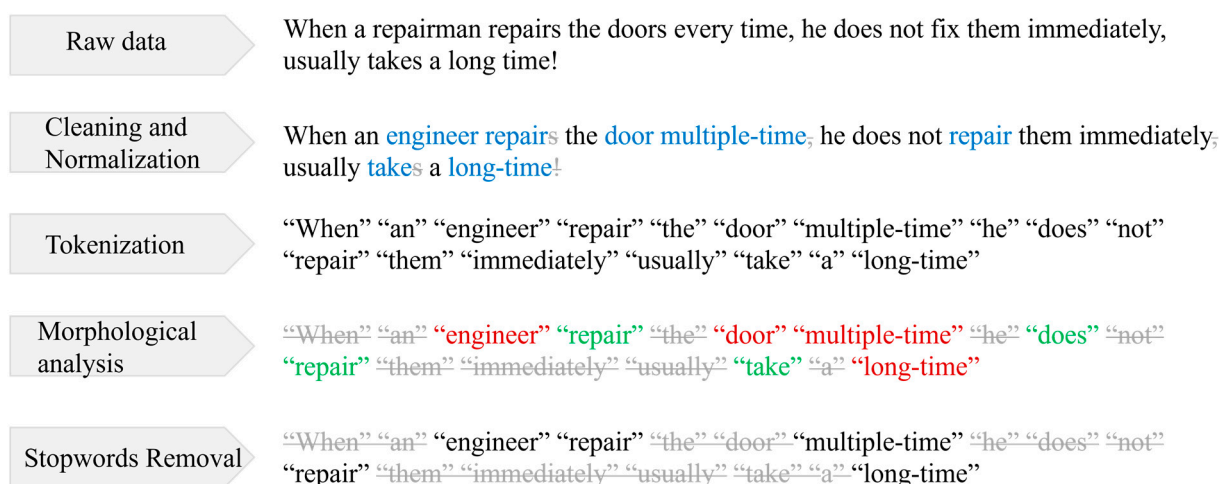
Table 1 lists the data collected in text format corresponding to the occupant ratings of “Dissatisfied” and “Very Dissatisfied”. A total of 151,272 and 76,837 space-separated words were identified in the subsets corresponding to the occupant ratings of “Dissatisfied” and “Very Dissatisfied”, respectively, with a given datapoint in the two subsets containing an average of 13.2 and 23.6 space-separated words, respectively. The collected data represent the following information: (i) the occupant’s sentiments toward the repair process (e.g., “I am frustrated” and “too bad”); (ii) the objects of the construction defects to be repaired (e.g., “doors in my unit” and “a wall in my master bedroom”); (iii) the nature/status of the repair process (“rip out and paint a wall”, “a lot of nails on the wall”, “drill a few holes in the bond”, and “not sticking to the wall”); and (iv) the number/frequency of repair service visits after filing a complaint to report the defect (e.g., “the third time”, “multiple times”, and “every time”). In addition, there are some general terms that represent concerns related to repair work (e.g., “correct” and “right way to do it”), although the terms have little relation to the dissatisfaction factors relating to the process of repairing construction defects.

Table 1. Examples of raw data.

Category	Raw Data
Dissatisfied	This is the third time I've had to file a claim to repair doors in my unit. I have a lot of problems with the repair service! When an engineer repairs the doors every time, he does not fix them immediately, usually takes a long time! It's too bad to repair the doors multiple times.
Very Dissatisfied	This is the third time I've had to rip out and paint a wall in my master bedroom. The repairman made a lot of nails on the wall, partially punctured it and injected bond. I am frustrated that the drywall is not sticking to the wall. I want to ask if it is correct to drive a few nails and drill a few holes in the bond. Is this the right way to do it?

3.2. Text Pre-Processing

The main objective of text pre-processing is to convert the collected data in unstructured text format into meaningful terms for SNA purposes [15,45,52]. Toward this objective, the text pre-processing in the present study consisted of cleaning and normalization, tokenization, morpheme analysis, and removal of stopwords. As shown in Figure 3, sources of data noise such as punctuation marks (e.g., “!”, “?”, and “,”) and index numbers were eliminated from the collected data, since they do not contain any meaningful information. Then, terms with potentially ambiguous or synonymous meanings were replaced with alternate terms. For example, “every time” and “long time” were converted to “multiple-time” and “long-time”, respectively, in order to distinguish between units of “time” (i.e., minutes, hours, and days) and number of “times” (i.e., occasions) that the occupant received repair services. Moreover, “repair” was used as the representative word for “fix”, “recover”, “restore”, and “reconstruct”. For normalization of the data, meanwhile, lemmatization was conducted. Lemmatization, it should be noted, is the process of grouping together different inflected forms of the same word. In other words, lemmatization returns the “lemmas” (or “root”) of the word. Taking the same example mentioned above, “repairs” was replaced by “repair” since “repair” is the root word of “repairs” (as well as of “repairing” and “repaired”, for that matter).

**Figure 3.** Examples of text pre-processing procedures.

The second step following cleaning/normalization was tokenization, whereby each sentence in the dataset was parsed into individual words. For example, the sentence “when an engineer repairs the doors multiple-times he does not repair them immediately usually takes a long-time” was split into 16 words: “when”, “an”, “engineer”, “repair”, “the”, “door”, “multiple-time”, “he”, “does”, “not”, “repair”, “them”, “immediately”, “usually”, “take”, “a” and “long-time”.

In the third step, morphological analysis was implemented to identify and extract important terms related to occupant dissatisfaction from the tokenized data. In general, in Korean, nouns are the words that contain the critical information regarding occupants' experiences and satisfaction/dissatisfaction [15]. However, adjectives (e.g., inadequate, unkind) and verbs (e.g., repair, promise) also contain some important information in this regard. As such, nouns, adjectives, and verbs were all retained as potential keywords at this juncture. To conduct the morphological analysis, the KoNLPy Python package was used, as this package has been widely used in applications involving the Korean language [50]. Particularly, Okt class in the KoNLPy package was used to normalize words and extract stems of words. Taking the same example sentence presented above, the nouns (red color in Figure 3) include four semantic words—"engineer", "door", "multiple-time", and "long-time"—while the verbs (green color in Figure 3) include "repair", "does", "repair", and "take".

The fourth and final step in text pre-processing was removal of stopwords, which eliminated from the analysis extremely common terms with little analysis value [51]. In text pre-processing, the list of stopwords is typically generated by sorting the most frequent terms and removing them manually from the collected data. The list of stopwords in this research includes common verbs (e.g., "take" and "does") and words related to objects or states of construction defects (e.g., "doors"). At this juncture, it should be noted that this paper mainly focuses on identifying factors affecting dissatisfaction for occupants living in newly constructed apartments in terms of the construction defect repair process instead of the objects and/or states of construction defects. In line with this goal, the focus of the analysis in this paper is streamlined towards the removal of construction defects, their associated objects, and states, ensuring a more efficient and precise mining analysis. Considering the same example sentence mentioned above, removal of stopwords results in five remaining semantic words: "repair", "multiple-times", "engineer", "repair", and "a-long-time".

3.3. Identification of Occupant Dissatisfaction Factors

3.3.1. Keyword Extraction

Keyword extraction plays a critical role in information extraction, text categorization, text classification, text summarization, and information retrieval since it identifies the most important words and features, which in turn become clues to understanding the text data. There are a number of different keyword extraction techniques used in text mining, such as TF, TF-IDF, and rapid automatic keyword extraction (RAKE). TF calculates the number of times a given word occurs in a document compared with the total number of words in the document. In other words, a word with a high TF count is considered more important than words with lower TF counts in the given dataset. Due to its straightforward and intuitive calculation process, many studies have adopted TF for extracting keywords from text data [12,15]. However, TF may not be a reliable method for measuring how important a term is within a text relative to the entire text dataset [53]. Given the objective of the present study, a TF-IDF method was adopted, where TF-IDF was calculated using Equation (1). TF-IDF is a traditional keyword extraction method that mainly assesses the relative importance of terms or phrases within the collected dataset based on TF and IDF parameters. The main principle underlying the concept of IDF is that a given word is considered not highly representative and of low importance within the collected dataset when its frequency is high. Based on application of these parameters, a word is considered important when its TF-IDF score is high.

$$TF - IDF_{a,b} = tf_{a,b} \times \log_2 \left(\frac{N}{df_a} \right) \quad (1)$$

where $TF - IDF_{a,b}$ = the TF-IDF weight of term a in comment b ; $tf_{a,b}$ = absolute frequency of term a in comment b ; N = the total number of comments in the collected data; df_a = the number of comments containing term a .

Visualization (i.e., building information modeling) is commonly used in the construction domain to monitor, control, and aid understanding of project progress. In information management, visualization is a supportive tool that aids decision makers by providing insights into complex concepts and identifying emerging or changing patterns [54]. In the present study, keywords are visualized using a tag cloud that presents in pictorial and graphical formats and makes use of different font sizes, colors, and distances to add further clarity. For example, higher TF-IDF scores of keywords are represented with larger font sizes.

3.3.2. Identification of Relationships among Keywords

SNA, a text-mining technique, is implemented to establish and recognize relationships between keywords extracted from textual data. SNA is increasingly being used in a range of different application areas, such as political discourse analysis, analyzing causes of project delays in construction, cognitive psychology, and human semantic memory [55–57]. It represents results in a manner that is intuitive and easy to understand by making use of various visualization techniques, using nodes and edges to represent keywords and relationships among keywords, respectively. In other words, SNA, also called “word network analysis”, can transform unstructured text data into a structured text network in order to discover relationships among keywords based on nodes and edges (also called “links”). SNA essentially develops a semantic network by calculating co-occurrence of keywords, where co-occurrence is defined as the appearance of words together in a sentence, paragraph, or text [58]. In this respect, the co-occurrence-based semantic network in the present study was built to recognize the relationships between the extracted keywords. A detailed explanation of the manner in which keyword co-occurrence is calculated is provided in a previous study [58].

To aid understanding of the characteristics of the semantic networks, SNA makes use of several evaluation metrics: degree centrality (DC), closeness centrality (CC), and betweenness centrality (BC). Centrality describes the location of a given node relative to the center of the entire network. In this context, DC reflects the number of nodes that a given node is directly connected to. In other words, a high value of DC means that the node is related to a large number of other nodes in the network [59]. The DC of the k th node is calculated using Equation (2):

$$DC(k) = \frac{E_k}{TE} \quad (2)$$

where E_k = the number of edges directly connected to the k th node and TE = total number of edges in the network. CC, meanwhile, indicates the degree of closeness of a given node to all other nodes in the network [60], where a high value of CC is indicative of a high degree of closeness to other nodes in the network. A high value of CC also means that the given node is likely to be sensitive to the effect of other nodes, and vice versa. The CC of the k th node can be calculated using Equation (3) (based on computation of the geodesic distance from the node to all the other nodes):

$$CC(k) = \frac{1}{\sum d_{k,l}} \quad l = 1, 2, \dots, n \quad (3)$$

where $CC(k)$ = closeness centrality of the k th node and $d_{k,l}$ = distance between node k and node l . BC, also referred to as “intermediation centrality”, measures the number of times a node appears as a bridge node between the paths connecting pairs of other nodes. In this respect, a node with a high BC as computed using Equation (4) is one that has a significant effect on the flow of information within the network:

$$BC(k) = \sum_{i,j} \left[\frac{P_{i,j}(k)}{P_{i,j}} \right] \quad (4)$$

where $BC(k)$ = betweenness centrality of k th node; P_{ij} = the number of all shortest paths between node i and node j ; and $P_{ij}(k)$ = the number of all shortest paths between node i and node j that pass through k th node.

Based on these SNA parameters, the influential keywords (i.e., the keywords with a high co-occurrence with other words) were identified as the basis for determining the occupant dissatisfaction factors during the construction defect repair period in apartment buildings.

4. Results

The text pre-processing was implemented with the collected dataset which consisted of subsets of 9058 and 3816 datapoints corresponding to the occupant ratings of “Dissatisfied” and “Very Dissatisfied”, respectively. In the process of text pre-processing for translating Korean to English, multiple iterations were employed to replace ambiguous and synonymous words. Initially, the research team manually grouped these words and substituted them with a single representative word, referencing a standard Korean language dictionary [61]. Subsequently, during the Korean-to-English translation, this replacement procedure was carried out using the WN Python package [62] and supplemented with manual adjustments based on the Collins thesaurus [63]. The WN package, including synsets, lemmas, hypernyms, and hyponyms classes, helped to aggregate and manage the similar and/or same meanings of words in the dataset, which were replaced by one representative word for efficient and effective data analysis in a text format. For example, “slow”, “late”, “postpone”, “delay”, and “defer” were all replaced with “delay” as the representative word. As further examples, “C/S”, “service”, “center”, and “CS” were replaced with “service”, and “receipt” was used to represent “file”, “inform”, “record”, “receive”, “claim”, and “declare”. Following synonym replacement, there were 28,251 words in the “Dissatisfied” subset and 5119 words in the “Very Dissatisfied” subset. However, these data still included unnecessary words that do not contain information that serves the objective of the present study, which is to identify the occupants’ dissatisfaction factors in newly constructed apartment buildings in terms of the construction defect repair process when residents claim defects in their units. There were 498 words associated with objects and states of defects (e.g., ‘defect’, ‘kitchen’, ‘room’, ‘door’, ‘window’, ‘plumbing’, etc.), while words related to company and occupant information (such as brand name, address, bank account number, name of customer service center, etc.) amounted to 10,015 and 18,222 respectively. These words were manually identified and removed. Table 2 presents some example results of the text pre-processing.

Table 2. Examples of text pre-processing results.

Category	Raw Data	Pre-Processed Data
Dissatisfied	Defects may occur but repeated defects are very inconvenient as an occupant. When defects are repaired again, please handle them meticulously to prevent repeated defects.	“repeat”, “inconvenient”, “occupant”, “repair”, “handle”, “again”, and “prevent”
Very Dissatisfied	It seems that there is no intention to file the defects. Also, inform wrong repair schedule. According to an informed day to repair the defect, I was at home but no one visited.	“wrong”, “intention”, “repair”, “visit”, “receipt”, “day”, “inform”, and “schedule”

Based on the results of the text pre-processing, TF-IDF was computed to extract keywords in both the “Dissatisfied” and “Very Dissatisfied” subsets. Figure 4a shows the top 30 keywords extracted from 9058 pre-processed data records in the “Dissatisfied” category. Among the top 30 keywords, “Repair” has the highest TF-IDF, followed by “Response”, “Receipt”, “Request”, “Prohibit”, and “Visit”. Moreover, the following patterns were identified in the keywords: (i) unprofessional conduct of receptionists and workers corresponds with words such as “Attitude”, “Unkindness”, “Disappointment”, “Response”, “Worker”,

and “Confirmation”; (ii) failure on the part of the contractor to keep an appointment to repair defects corresponds with the words “Visit”, “Promise”, “Again”, “Delay”, “Contact”, and “Change”; and (iii) inadequate repair work corresponds with the words “Again”, “Deficient”, “Accuracy”, “Clean-up”, and “Finish”. Figure 4b includes a word cloud visualizing the top 30 keywords in the “Dissatisfied” subset.

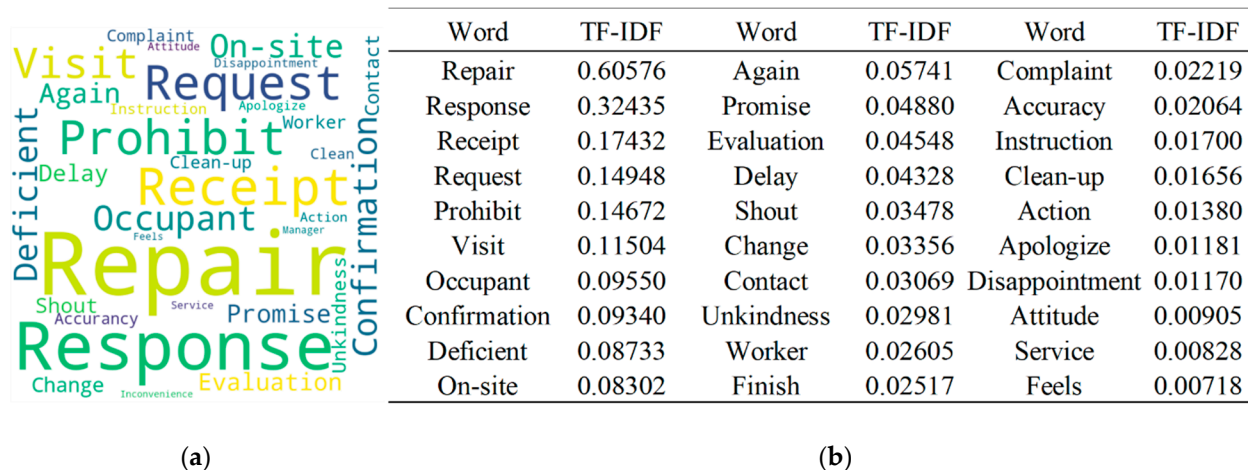


Figure 4. (a) Word clouds with top 30 keywords; (b) top 30 keywords in the “Dissatisfied” subset of the data.

Figure 5a lists the top 30 keywords extracted from the 3816 records in the “Very Dissatisfied” subset of the data. According to the results of the TF-IDF calculation, the top five keywords are “Repair”, “Complaint”, “Visit”, “Promise”, and “Receipt”. Moreover, based on the TF-IDF results for both the “Dissatisfied” and “Very Dissatisfied” categories, the occupants reported the most negative experiences using words such as “Accuracy”, “Again”, “Time”, “Problem”, “Inconvenience”, and “Exchange”. The most frequent complaints among occupants assigning a rating of “Dissatisfied” were wasted time (due to the contractor failing to keep the repair schedule) and multiple visits to the occupant’s unit. In both the “Dissatisfied” and “Very Dissatisfied” categories, there are a total of nine common words corresponding with occupant dissatisfaction factors: “Repair”, “Visit”, “Promise”, “Receipt”, “Unkindness”, “Again”, “Delay”, “Accuracy”, and “Clean-up”. Based on these words, the dissatisfaction factors can be broadly classified into three categories: (i) inadequate repair work (e.g., “Clean-up”, “Mess”, and “Problem”), (ii) disorganized repair schedule (e.g., “Time”, “Delay”, “Promise”, and “Schedule”) due to lack of communication between the repair service center and occupants, and (iii) conduct of receptionists (e.g., “Unkindness”, “Response”, “Attitude”, and “Receipt”). Figure 5b includes a word cloud visualizing the top 30 keywords in the “Very Dissatisfied” subset.

To identify the relationships among the extracted keywords, SNA was used to develop word networks based on the top 50 word relationships (i.e., co-occurrence). In the SNA results, the word network in the “Dissatisfied” category has 42 nodes, 859 edges, a density of 0.808, and a total of 102,688 co-occurrences. In the “Very Dissatisfied” subset of the data, there are 45 nodes, 861 edges, a density of 0.769, and 56,262 co-occurrences. Table 3 provides examples of word relationships acquired from the word networks for “Dissatisfied” and “Very Dissatisfied” based on these results. As can be seen, in both the “Dissatisfied” and “Very Dissatisfied” subsets of the data, “Repair + Inadequate” has the highest co-occurrence at 683 and 608, respectively, meaning that “Repair” and “Inadequate” appear together in 683 “Dissatisfied” records and 608 “Very Dissatisfied” records. Regarding the word relationships within the “Dissatisfied” subset of the data, “Repair + Fulfillment” (565 co-occurrences), “Repair + Promise” (461 co-occurrences), “Repair + Visit” (396 co-occurrences), “Repair + Implementation” (385 co-occurrences), and “Repair + Delay” (350 co-occurrences) are the highest co-occurrences. Regarding the word relationships

within the “Very Dissatisfied” subset of the data, “Repair + Delay”, “Repair + Receipt”, “Repair + Again”, “Repair + Complaint”, “Repair + Accuracy”, and “Occupant + Promise” have 368, 270, 159, 118, 111, and 106 co-occurrences, respectively. It is crucial to underscore that the authors have chosen the top 10 word relationships because of the limited occurrence of co-occurrences considering the size of the dataset used in this paper.

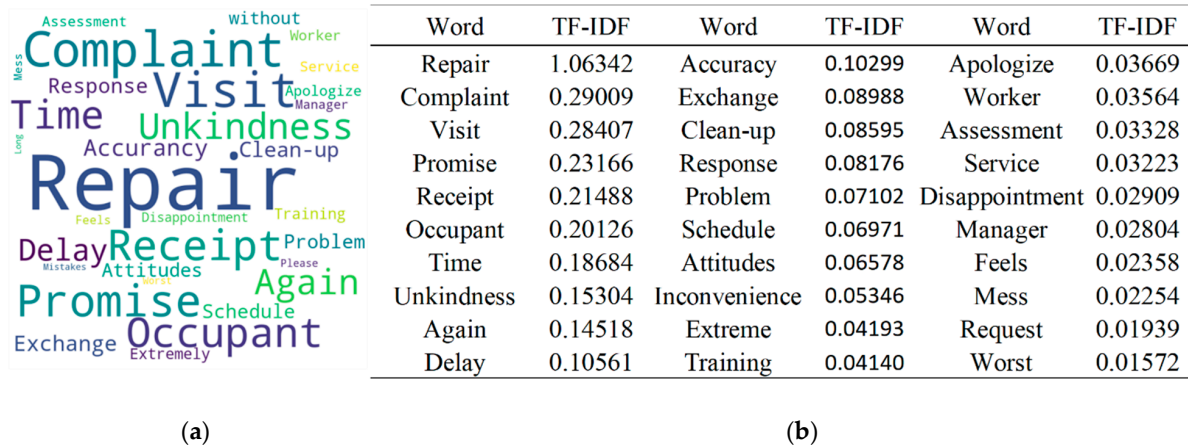


Figure 5. (a) Word clouds with top 30 keywords; (b) Top 30 keywords in “Very dissatisfied” subset of the data.

Table 3. Examples of word relationships in the word networks for both the “Dissatisfied” and “Very Dissatisfied” subsets of the data.

Category	Word Relationship	Co-Occurrence
Dissatisfied	Repair + Inadequate	683
	Promise + Fulfillment	565
	Repair + Promise	461
	Repair + Visit	396
	Repair + Implementation	385
	Repair + Delay	350
	Repair + Deficient	315
	Occupant + Promise	271
	Unkindness + Response	200
Very Dissatisfied	Attitude + Unkindness	181
	Repair + Inadequate	608
	Repair + Delay	368
	Repair + Receipt	270
	Repair + Again	159
	Repair + Complaint	118
	Repair + Accuracy	111
	Occupant + Promise	106
	Response + Unkindness	99
	Again + Visit	85
	Visit + Inadequate	78

Centrality measures, expressed in terms of DC, CC, and BC, are concepts which are commonly used in SNA in order to understand the significance of words or concepts within a body of text or a knowledge graph. By calculating and analyzing centrality scores, key terms, concepts, or words that are central to the overall meaning or structure of the semantic network can be identified. As described above, DC indicates that words are commonly used or play a significant role in connecting other nodes, CC represents the degree of closeness to all other nodes, and BC represents the magnitudes of bridge words between the paths connecting the pairs of other words in the word networks. Table 4 lists the top 16 words in the word networks in terms of DC, BC, and CC for the “Dissatisfied”

and “Very Dissatisfied” subsets of the data. As can be seen, “Repair” has the highest values of DC, BC, and CC in both the “Dissatisfied” subset (0.641, 0.107, and 0.735, respectively) and in the “Very Dissatisfied” subset (0.787, 0.094, and 0.824, respectively). Based on these values, “Repair” under “Very Dissatisfied” is more a central word with higher closeness than the ones under “Dissatisfied”, and it also plays a more important function as a bridge word linking to the words under “Dissatisfied” than do the other words under “Very Dissatisfied”. The DC value of “Response” in the “Dissatisfied” word network is 0.553 and its CC and BC are 0.69 and 0.066, respectively. The keyword “Clean-up” has the smallest DC value (0.203), indicating that it has little contact with other nodes in the word network. Its BC and CC values are also relatively small at 0.004 and 0.555, respectively, meaning that it is far from the center of the network. The DC and CC values of “Inadequate”, meanwhile, are 0.491 and 0.662, respectively. Its BC value, at 0.020, is slightly smaller than that of “Receipt”, meaning that “Receipt” affects other keywords more than does “Inadequate”. In the “Very Dissatisfied” word network, “Visit” was also identified as an influential factor, with a DC value of 0.522, a BC of 0.022, and a CC of 0.676. “Fulfillment”, “Inadequate”, and “Unkindness” were found to have similar degrees of influences in the network since their BC values are the same and their respective DC and CC values are not significantly different. “Response” has the smallest DC value at 0.316, indicating that it has little contact with the other nodes in the network. Its CC and BC values are also relatively small at 0.593 and 0.005, respectively, meaning that “Response” is the weakest factor in the network.

Table 4. Top 16 words and their DC, BC, and CC in the word networks for “Dissatisfied” and “Very Dissatisfied”.

Category	Word	DC	BC	CC
Dissatisfied	Repair	0.641	0.107	0.735
	Response	0.553	0.066	0.690
	Inadequate	0.491	0.020	0.662
	Receipt	0.385	0.021	0.618
	Visit	0.365	0.017	0.611
	Occupant	0.362	0.015	0.609
	Complaint	0.313	0.015	0.591
	Time	0.279	0.010	0.580
	Accuracy	0.260	0.007	0.573
	Fulfillment	0.233	0.006	0.565
	Unkindness	0.228	0.006	0.562
	Again	0.214	0.005	0.558
	Promise	0.211	0.005	0.557
	Attitude	0.210	0.005	0.557
	Delay	0.208	0.005	0.556
	Clean-up	0.203	0.004	0.555
Very Dissatisfied	Repair	0.787	0.094	0.824
	Visit	0.522	0.022	0.676
	Promise	0.474	0.018	0.655
	Again	0.445	0.013	0.643
	Time	0.431	0.012	0.637
	Fulfillment	0.422	0.011	0.633
	Inadequate	0.418	0.011	0.632
	Unkindness	0.418	0.011	0.632
	Complaint	0.407	0.010	0.627
	Receipt	0.389	0.009	0.620
	Occupant	0.384	0.009	0.618
	Accuracy	0.384	0.009	0.618
	Service	0.378	0.008	0.616
	Change	0.350	0.007	0.606
	Delay	0.341	0.006	0.603
	Clean-up	0.325	0.006	0.597
	Response	0.316	0.005	0.593

In Figure 6, the word networks for “Dissatisfied” and “Very Dissatisfied” are visualized on 2D maps. Figure 6a represents the word network for “Dissatisfied”. This network has two keywords (red color)—“Repair” and “Response”—located in the center of the network. According to the words connected to the central word, “Repair”, occupants have had experiences involving the terms “Inadequate”, “Promise”, “Fulfillment”, “Receipt”, and “Delay”. In other words, the occupants have encountered issues related to improper repair work, difficulty reporting defects to the repair service center, and failure on the part of the contractor to keep the repair schedule. The other central word, “Response”, has relationships with words such as “Delay”, “Unkindness”, and “Complaint”. In other words, the occupants have encountered issues related to unkind responses from the representatives in the repair service center and slow repair work (i.e., time delays). The words that have relationships with the center words in the network are represented in green color in the figure.

Figure 6b illustrates the word network for “Very Dissatisfied”. This network has two central words: “Repair” and “Visit”. The central word, “Repair”, has relationships with “Unkindness”, “Promise”, “Receipt”, “Fulfillment”, “Delay”, “Inadequate”, and “Accuracy”. Based on these relationships, we can infer that occupants have encountered issues with repair work being inaccurate and delayed and with receptionists in the center acting in an unprofessional manner in fielding concerns about unfulfillment of repair work. The terms “Delay”, “Again”, “Promise”, “Time”, and “Change” are connected to the other central keyword, “Visit”. From this, we can infer that occupants have encountered issues related to (i) repair schedules being changed frequently without the changes being communicated to occupants and (ii) repairs being conducted in multiple iterations due to the defects not having been fixed correctly in the first place.

Based on the results of SNA and TF-IDF, we infer that the occupant dissatisfaction factors are: (i) inaccurate and inadequate repair work leading to wasted time (multiple visits), discomfort, and inconvenience for occupants, reflected in words such as “Repair”, “Visit”, “Accuracy”, “Inadequate”, and “Again”; (ii) failure to adhere to the agreed-upon terms for the repair work (e.g., “Fulfillment”, “Promise”, “Change”, and “Delay”); and (iii) unprofessional conduct from representatives in the repair service center (e.g., “Response”, “Attitude”, and “Receipt”). These findings have been shared with the collaborating company, and the comments examined in detail to better understand the causes underlying these factors. At this juncture, it should be noted that this paper presents a few examples of each dissatisfaction factor, but does not propose strategies to address the identified dissatisfaction factors (since this is not the objective of the present study). It is concluded that inaccurate and inadequate repair work is the result of a failure to satisfy occupants’ requirements (e.g., clean-up after the repair work). In addition, due to incorrect diagnosis of the defects, the original defects reported by the occupants often go unaddressed. As a result, repair of the defects takes an unnecessarily long time, with multiple visits leading to inconvenience and discomfort for occupants. As for the failure to adhere to the agreed-upon terms of the repair work, there are two main reasons for this. First, the engineers or contractors repairing the defects often fail to provide sufficient notice when visiting the unit to complete the repairs. Second, the repair schedule is frequently changed and appointments postponed without consideration of the occupant’s schedule. Tactlessness typically takes the form of inappropriate attitudes/conduct of the representatives in the service center (e.g., failing to properly record the defects) when occupants report defects or inquire about the status of a repair (e.g., repair schedule).

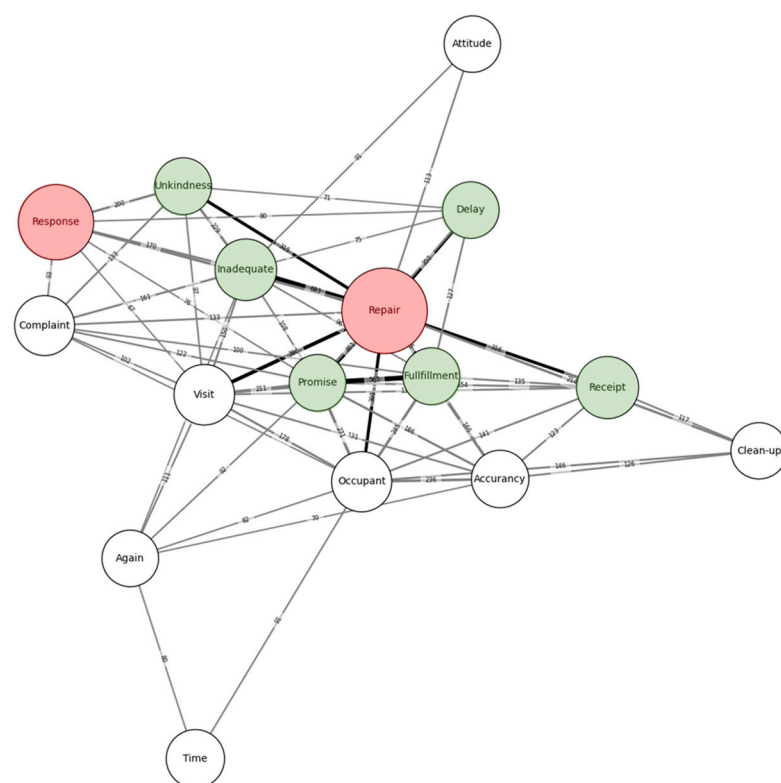
Table 5 illustrates the term frequency (TF) of construction defects associated with the previously identified factors, categorized under “dissatisfaction” and “strong dissatisfaction.” On the whole, occupants expressed significant dissatisfaction, particularly concerning telecommunications-related defects during the construction defect repair process. This dissatisfaction stems from factors such as inaccurate and inadequate repair work (factor 1), failure to meet requirements (factor 2), and unprofessional conduct (factor 3), all of which exhibit the highest TFs in both “dissatisfaction” and “strong dissatisfaction.” However, it

is noteworthy that in the context of “dissatisfaction,” factors 2 and 3 do not hold the top rank. When considering “dissatisfaction,” factor 1, which includes issues with windows, washrooms, and electricity-related defects, is ranked second, third, and fourth, respectively. Moreover, in the case of electricity and internet-related defects, occupants’ requirements were not fully met by the construction company, further contributing to their dissatisfaction. In the context of “strong dissatisfaction,” occupants expressed extreme displeasure primarily concerning inaccurate and inadequate repair work, which is associated with factor 2, affecting doors, electricity, and washroom-related defects. As a result, this information provides valuable insights for the company, indicating where improvements are needed within the construction defect repair services for occupants residing in newly constructed apartments.

Table 5. TFs of construction defects associated with occupant factors of dissatisfaction and strong dissatisfaction.

Defects	Dissatisfaction			Strong Dissatisfaction		
	Factor 1	Factor 2	Factor 3	Factor 1	Factor 2	Factor 3
Windows	9952	7910	6676	8186	4626	4724
Wallpaper	6221	4803	4037	5889	3245	2695
Doors	6387	6077	6007	13,958	7755	7827
Kitchen	6331	4562	4189	6982	3538	2700
Waterproof	9019	6244	2081	9019	6938	6938
Boiler	3656	2825	2908	4404	2742	3490
Washroom	9750	8125	2708	9208	7042	5958
Faucets and sanitary facilities	4877	3405	3466	4509	2209	2515
Air conditioning pipes	3297	2839	2106	4854	3022	4121
Floor	6076	4181	3371	3685	2064	1895
Internet	9208	9208	13,812	4604	4604	0
Furniture	8781	7064	6943	9105	5207	5274
Lighting	3123	2970	3154	3184	1868	2664
Ceiling	5995	4961	4134	4134	2274	2687
Electricity	9520	9587	8364	9413	6059	7472
Tile	6031	4845	3583	4668	2599	2624
Range hood	4577	3651	3215	4086	2234	3160
Telecommunication	12,244	7792	7792	14,470	11,131	13,357
Painting	5651	4416	4226	6458	4179	3941

Instead of verification of the results applying to the defect repair process in practice, this study included consultation with experts in the collaborating company. During the consultation, the company stated “We were aware that our defect repair process did not fully satisfy the occupants’ requirements, but we were uncertain about the specific factors causing occupant dissatisfaction. Additionally, we were surprised to discover instances of unprofessional conduct among our representatives in the repair service center, despite our regular training based on our manual, which outlines how to respond to occupants when they report defects.” As a result, this study identifies target factors that require the company’s attention to improve the serviceability and quality of the defect repair process.



(a)



(b)

Figure 6. Word network maps: (a) word network for “Dissatisfied”; (b) word network for “Very Dissatisfied”.

5. Discussion and Future Directions

The research team had meetings with the industrial partner to share the results of the proposed methodology involving the list of keywords for each of identified dissatisfaction factors. Based on the dissatisfaction factors with keywords, the company and research team investigated problems in the construction defect repair process, which included: (i) carelessness in the repair works since laborers focused on repairing only the defects and overlooked other aspects (e.g., cleaning), considering these as customers' responsibilities; (ii) significant gaps in the level of completion of repairing the defects between the laborers and customers, since customers generally want repair of the defects to be completed at once but the laborers consider multiple works if customers are not satisfied; (iii) number of completed tasks required to repair defects in a day, since the laborers should complete a certain number of repairs of defects per day as defined by their companies (sub-contractors). Due to this reason, the laborers generally reported cases which involved the delay of the repair date as completion of repairing the defects; (iv) lack of explanation to occupants before/after repairing the defects; (v) failure to fulfill customer appointments without notice in advance; (vi) lack of repair priorities (occupancy vs. non-occupancy in units). This lack may lead to the laborers repairing defects in non-occupied units, which do not urgently need repair of the defect, while the occupied units require repair of defects in a timely manner to improve the level of residents' satisfaction; (vii) lack of tracking and monitoring the status of repairing the defects; (viii) lack of communication and information sharing between the industrial partner and their sub-contractors; and (ix) lack of professional training tools and a standard response manual for representatives in the customer service center. At this junction, it should be noted that the combinations of keywords for the identified dissatisfaction factors were used as reference to identify the problems in the construction defect repair process. To address the problems, as shown in Figure 7, the industrial partner and research team have developed the following plans: (i) developing a management system for not only the repair laborers but also sub-contractors, called RLSC in this paper; (ii) development of an automated text message system to provide notice of the repair schedule (e.g., sending text messages one day and two hours before the visit date and time) and status of the repair works to occupants in an efficient manner; and (iii) development of standard response manuals which will be used regularly to train representatives in the customer service center. The aims of RLSC are to manage and record the history of repair works for the laborers and sub-contractors, such as the number of complaints from customers and second or more defect occurrences due to incompleteness of repairing the defects, and to select excellent laborers and sub-contractors depending on the historical records involving the level of occupants' satisfaction and number of defect occurrences after completion of repair works. Furthermore, these historical data will be used to identify unqualified laborers and evaluate the performance of the sub-contractors. The unqualified laborers will not be hired and the results of performance evaluation for sub-contractors will be used as one of the criteria to contract them in future projects. In future, the effectiveness of the proposed systems will be validated with sufficient occupant complaints.

Although RLSC, the automated text message system, and standard response manuals are proposed to solve most of challenges described above, the limitations (vi), (vii) and (viii) have not been fully addressed yet. In line with limitation (viii), lack of communication and information sharing, one of the residents stated *"When I try to ask about the status of repairs, the receptionist tells me that they are not authorized to do so, and there is no direct line to a person in charge of each job. The whole system of handling repairs makes really difficult for residents to know when defects are going to be fixed or what's going on"*. To address this limitation, blockchain technology has recently gained high attention as an information sharing platform since it provides traceable, reliable, secured, and transparent data through the electronic ledger of digital information and a group of consensus protocols for secured and reliability [64]. Consequently, project participants within the blockchain network have the ability to cooperate in the tasks of recording, validating, storing, and retrieving

repair-related information. In terms of tracking and monitoring the repair works efficiently in the blockchain network, advanced technologies such as 3D scanners and high-resolution digital images from advanced cameras can be applied to track, monitor, and store the status of repair works and of defects efficiently and effectively in the blockchain platform. In addition, these digital data can be used as resources to execute objectively the inspection to confirm the defects using various computer vision algorithms such as convolutional neural networks (CNNs) and support vector machine-based recognition [65,66]. To address the lack of repair priorities, the workflow of the construction defect repair process, particularly the scheduling aspect, should be optimized based on application of optimization algorithms or simulation [67,68]. Although the standard response manual is proposed to train the professional attitudes of representatives, a reward system for representatives in the center may be considered since it will encourage workers have more responsibility and motivate their work professionally [69].

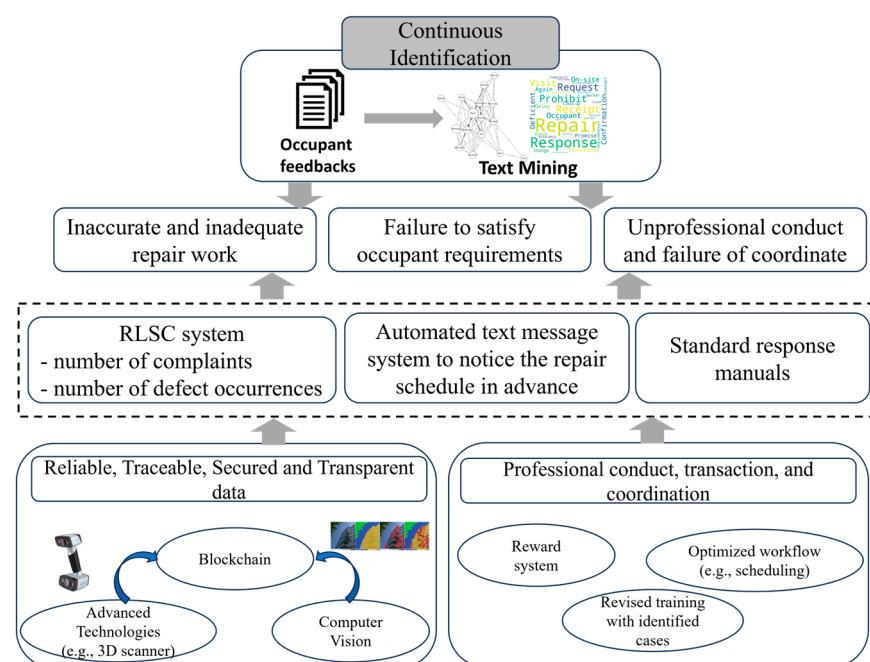


Figure 7. A future framework for quality and serviceability improvements.

Further opportunities exist for improvements to enhance this analysis. Although the proposed methodology as a pilot study used a dataset of 12,874 comments, the volume of the dataset may not be sufficient to generalize occupants' dissatisfaction factors identified by the proposed methodology. Therefore, the proposed methodology should be implemented to identify occupants' dissatisfaction factors continuously with more volume of occupant feedback. Occupants may have various interests and concerns over time and across different geographical areas in South Korea. In this respect, popularity analysis can be used to identify interests and/or concerns in the construction defect repair services over time and across geographical areas in South Korea. The results of popularity analysis will provide insights to the managers and research team regarding where and how to enhance or improve particular aspects of the construction defect repair process in order to satisfy the occupants' requirements. Furthermore, there may be a relationship between the occupants' dissatisfaction factors identified in this paper and apartment features such as total floor areas of the apartment units and types of construction defects. Verification of the proposed methodology should be completed by using other text-mining techniques. For example, topic modeling (i.e., latent Dirichlet allocation) may be useful to acquire major topics of occupants' complaints from a large volume of data consisting of various levels of occupants' satisfaction and dissatisfaction since these topics involve the list of occupants' experience factors. The results of the topic modeling and SNA can be compared to find out the most

efficient text-mining technique to identify the occupants' experience factors in the process of repairing the defects in newly constructed apartments. Once this comparison is completed, an automated system to identify the occupants' dissatisfaction factors from a large volume of data can be developed using the best text-mining technique and applied into other newly constructed facilities (e.g., commercial buildings).

6. Conclusions

With apartment buildings being constructed at an increasing rate in South Korea, construction defects being encountered in the occupancy stage has emerged as a serious social issue, with these defects sometimes giving rise to disputes between occupants living in newly constructed apartments and general contractors. To mitigate this issue, the aim of this study is to not only understand occupant dissatisfaction factors during the construction defect repair process in newly constructed apartments but also identify problems in the construction defect repair process. The results of this analysis as resources will provide information regarding where and how to enhance and improve the quality and serviceability of construction defect repair services. To the authors' knowledge, this study is the first to use text-mining techniques to analyze occupant complaints (9058 records in the "Dissatisfied" subset and 3816 records in the "Very Dissatisfied" subset) filed following completion of repair work to correct defects in newly constructed apartments. Text pre-processing was implemented by replacing synonyms with representative words and removing the words associated with the objects and states of the defects and company and occupant information. The pre-processing data were used to extract keywords (e.g., "Repair", "Response", and "Prohibit") from the subsets of data corresponding to the occupant ratings of "Dissatisfied" and "Very Dissatisfied" based on TF-IDF calculation. Then, SNA was implemented to explore the word relationships and visualize them in the word network maps. In summary, this paper identifies the following occupant dissatisfaction factors: (i) inaccurate and inadequate repair work (e.g., "Repair", "Visit", "Accuracy", "Inadequate", and "Again"); (ii) failure to adhere to the agreed-upon terms for the repair work (e.g., "Fulfillment", "Promise", "Change", and "Delay"); and (iii) unprofessional conduct of the representatives in the repair service center (e.g., "Response", "Attitude", and "Receipt").

In view of its contributions, the method in this paper is proposed as a first study for identifying occupant dissatisfaction factors in the construction defect repair process in newly-constructed apartments by applying text-mining techniques to complaints written by occupants. Second, the identified dissatisfaction factors associated with the keywords can be used as resource to identify problems in the construction defect repair process efficiently and effectively. Based on the dissatisfaction factors and the keywords, the research team with the industrial partner have investigated and identified nine main problems in the construction defect repair services, including lack of repair priorities, significant gaps in the level of completion of repair works between the laborers and customers, failure to provide notice of changes to the repair schedule to occupants in advance, and lack of communication and information sharing between project participants and occupants. Third, the identification of problems will allow the company or other users to determine where and how to improve the serviceability and quality of the construction defect repair process. As a practical plan to address the limitations, the research team and the industrial partner will develop three main components, which are the RLSC system, an automated text message system, and standard response manuals. The RLSC system is to not only manage the history of repair works for the laborers and sub-contractors but also identify excellent laborers and sub-contractors depending on the historical records involving the level of occupants' satisfaction and number of defect occurrences after completion of the repair works. In addition, the historical data in the RLSC system will be used to identify unqualified laborers and evaluate the performance of sub-contractors. As a result, unqualified laborers will not be hired and the results of performance evaluation for sub-contractors will be used as one of the criteria to contract them in future projects. However, as described in the discussion section, the proposed methodology as a pilot study requires

continuous implementation with more volume of occupant complaints to generalize the identified occupant dissatisfaction factors.

Author Contributions: Conceptualization, S.M. and S.H.; Data curation, I.J. and S.-H.N.; Funding acquisition, S.M.; Investigation, I.J., S.H. and S.M.; Methodology, I.J. and S.H.; Supervision, S.M., S.H. and J.-J.K.; Writing—original draft, S.-H.N. and S.H.; Writing—review & editing: S.M. and J.-J.K. All authors have read and agreed to the published version of the manuscript.

Funding: 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).

Data Availability Statement: The data that support the findings of this study are available from the corresponding author upon reasonable request. The data are not publicly available due to privacy.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Construction and Economy Research Institute of Korea. Construction Industry Image Status and Improvement Plan. Available online: <http://www.cerik.re.kr/report/issue/detail/2457> (accessed on 9 March 2023).
2. Defect Examination Dispute Resolution Committee (DEDRC). Defect Inspection Dispute Settlement Casebook. Available online: <https://www.adc.go.kr/> (accessed on 9 March 2023).
3. Brogan, E.; McConnell, W.; Clevenger, C.M. Emerging Patterns in Construction Defect Litigation: Survey of Construction Cases. *J. Leg. Aff. Disput. Resolut. Eng. Constr.* **2018**, *10*, 03718003. [CrossRef]
4. Noble-Allgire, A. Notice and opportunity to repair construction defects: An imperfect response to a perfect storm. *Real Prop. Trust. Estate Law J.* **2009**, *43*, 729–796.
5. Grosskopf, K.R.; Oppenheim, P.; Brennan, T. Preventing defect claims in hot, humid climates. *ASHRAE J.* **2008**, *50*, 40.
6. Seo, H.S. Analysis of Risk Factors and Management Measures through Case Analysis of Apartment House Defect Disputes. Ph.D. Thesis, Yeungnam University, Gyeongsan, Republic of Korea, 2013.
7. Macarulla, M.; Forcada, N.; Casals, M.; Gangolells, M.; Fuertes, A.; Roca, X. Standardizing Housing Defects: Classification, Validation, and Benefits. *J. Constr. Eng. Manag.* **2013**, *139*, 968–976. [CrossRef]
8. Fauzi, S.N.F.; Yusof, N.; Abidin, Z.Z. The relationship of housing defects, occupants' satisfaction and loyalty behavior in build-ten-sellhouses. *Procedia Soc. Behav. Sci.* **2012**, *62*, 75–86. [CrossRef]
9. Au-Yong, C.P.; Azmi, N.F.; Mahassan, N.A. Maintenance of lift systems affecting resident satisfaction in low-cost high-rise residential buildings. *J. Facil. Manag.* **2018**, *16*, 17–25. [CrossRef]
10. Jiboye, A.D. Post-Occupancy evaluation of residential satisfaction in Lagos, Nigeria: Feedback for residential improvement. *Front. Archit. Res.* **2012**, *1*, 236–243. [CrossRef]
11. Million, R.N.; Alves, T.D.C.L.; Paliari, J.C. Impacts of residential construction defects on customer satisfaction. *Int. J. Build. Pathol. Adapt.* **2017**, *35*, 218–232. [CrossRef]
12. Villeneuve, H.; O'Brien, W. Listen to the guests: Text-mining Airbnb reviews to explore indoor environmental quality. *Build. Environ.* **2020**, *169*, 106555. [CrossRef]
13. Teng, J.; Guo, C.-P.; Li, J.-L.; Chen, Y.-Q.; Yang, X.-Z. The method of analyzing metro complaint data and its application. In Proceedings of the 17th COT International Conference of Transportation Professionals, Shanghai, China, 7–9 July 2017; pp. 2005–2016.
14. Drake, K.; Zechman, E.M. Using consumer complaints to characterize contamination events in a water distribution system. In Proceedings of the World Environmental and Water Resources Congress, Albuquerque, NM, USA, 20–24 May 2012; ASCE: Reston, VA, USA, 2012; pp. 2247–2252.
15. Chang, T.; Chi, S.; Im, S.-B. Understanding User Experience and Satisfaction with Urban Infrastructure through Text Mining of Civil Complaint Data. *J. Constr. Eng. Manag.* **2021**, *148*, 04022061. [CrossRef]
16. Assaf, S.; Srouf, I. Using a data driven neural network approach to forecast building occupant complaints. *Build. Environ.* **2021**, *200*, 107972. [CrossRef]
17. VanDemark, L.; Clevenger, C.M.; Click, M. Building Envelope Issues within Construction-Defect Litigation. *J. Leg. Aff. Disput. Resolut. Eng. Constr.* **2021**, *13*, 03721003. [CrossRef]
18. VanDemark, L.; Clevenger, C.M. Common disputes in eminent domain cases. *J. Leg. Aff. Disput. Resolut. Eng. Constr.* **2020**, *12*, 05019010. [CrossRef]
19. VanDemark, L.; Clevenger, C.M.; Brogan, E. Designating responsible parties for drainage within 1.5 m of a building. *J. Leg. Aff. Disput. Resolut. Eng. Constr.* **2019**, *11*, 03719001. [CrossRef]
20. Paton-Cole, V.; Aibinu, A.A. Construction defects and disputes in low-rise residential buildings. *J. Leg. Aff. Disput. Resolut. Eng. Constr.* **2021**, *13*, 05020016. [CrossRef]

21. Chong, W.-K.; Low, S.-P. Assessment of Defects at Construction and Occupancy Stages. *J. Perform. Constr. Facil.* **2005**, *19*, 283–289. [\[CrossRef\]](#)
22. Bazzan, J.; Echeveste, M.E.; Formoso, C.T.; Altenbernd, B.A.; Barbian, M.H. An Information Management Model for Addressing Residents' Complaints through Artificial Intelligence Techniques. *Buildings* **2023**, *13*, 737. [\[CrossRef\]](#)
23. Ma, N.; Zhang, Q.; Murai, F.; Braham, W.W.; Samuelson, H.W. Learning building occupants' indoor environmental quality complaints and dissatisfaction from text-mining Booking.com reviews in the United States. *Build. Environ.* **2023**, *237*, 110319. [\[CrossRef\]](#)
24. Sun, Y.; Kojima, S.; Nakaohkubo, K.; Zhao, J.; Ni, S. Analysis and evaluation of indoor environment, occupant satisfaction, and energy consumption in general hospital in China. *Buildings* **2023**, *13*, 1675. [\[CrossRef\]](#)
25. Kim, Y.K.; Abdou, Y.; Abdou, A.; Altan, H. Indoor environmental quality assessment and occupant satisfaction: A post-occupancy evaluation of a UAE University office building. *Buildings* **2022**, *12*, 986. [\[CrossRef\]](#)
26. Roumi, S.; Zhang, F.; Stewart, R.A.; Santamouris, M. Weighting of indoor environment quality parameters for occupant satisfaction and energy efficiency. *Build. Environ.* **2023**, *228*, 109898. [\[CrossRef\]](#)
27. Abdul-Rahman, H.C.; Wang, C.; Kamaruzzaman, S.N.; Mohd-Rahim, F.A.; Mohd-Danuri, M.S.; Lee, K. Case study of facility performance and user requirements in the University of Malaya research and development building. *J. Perform. Constr. Facil.* **2015**, *29*, 04014131. [\[CrossRef\]](#)
28. Baker, H.; Hallowell, M.R.; Tixier, A.J.P. Automatically learning construction injury precursors from text. *Autom. Constr.* **2020**, *118*, 103145. [\[CrossRef\]](#)
29. Zhang, F.; Fleyeh, H.; Wang, X.; Lu, M. Construction site accident analysis using text mining and natural language processing techniques. *Autom. Constr.* **2019**, *99*, 238–248. [\[CrossRef\]](#)
30. Sun, A.; Lachanski, M.; Fabozzi, F.J. Trade the tweet: Social media TM and sparse matrix factorization for stock market prediction. *Int. Rev. Financ. Anal.* **2016**, *48*, 272–281. [\[CrossRef\]](#)
31. Jung, H.; Lee, B.G. Research trends in text mining: Semantic network and main path analysis of selected journals. *Expert Syst. Appl.* **2020**, *162*, 113851. [\[CrossRef\]](#)
32. Meaney, C.; Moineddin, R.; Voruganti, T.; O'Brien, M.A.; Krueger, P.; Sullivan, F. TM describes the use of statistical and epidemiological methods in published medical research. *J. Clin. Epidemiol.* **2016**, *74*, 124–132. [\[CrossRef\]](#)
33. Al Qady, M.; Kandil, A. Automatic clustering of construction project documents based on textual similarity. *Autom. Constr.* **2014**, *42*, 36–49. [\[CrossRef\]](#)
34. Cheng, M.-Y.; Kusoemo, D.; Gosno, R.A. Text mining-based construction site accident classification using hybrid supervised machine learning. *Autom. Constr.* **2020**, *118*, 103265. [\[CrossRef\]](#)
35. Yu, W.-D.; Hsu, J.-Y. Content-based text mining technique for retrieval of CAD documents. *Autom. Constr.* **2013**, *31*, 65–74. [\[CrossRef\]](#)
36. Akanbi, T.; Zhang, J. Design information extraction from construction specifications to support cost estimation. *Autom. Constr.* **2021**, *131*, 103835. [\[CrossRef\]](#)
37. Zhou, Z.; Zhou, X.; Qian, L. Online public opinion analysis on infrastructure megaprojects: Towards an analytical framework. *J. Manag. Eng.* **2021**, *37*, 04020105. [\[CrossRef\]](#)
38. Zhong, B.; Xing, X.; Love, P.; Wang, X.; Luo, H. Convolutional neural network: Deep learning-based classification of building quality problems. *Adv. Eng. Inform.* **2019**, *40*, 46–57. [\[CrossRef\]](#)
39. Xie, P.; Xing, E.P. Integrating document clustering and topic modeling. In Proceedings of the 29th Conference UAI 2013: Uncertainty in Artificial Intelligence, Bellevue, DC, USA, 11–15 July 2013; Cornell University: Ithaca, NY, USA, 2013; pp. 694–703.
40. Blei, D.; Carin, L.; Dunson, D. Probabilistic topic models. *IEEE Signal Process Mag.* **2010**, *27*, 55–65. [\[CrossRef\]](#) [\[PubMed\]](#)
41. Wu, K.; Zhang, J.; Huang, Y.; Wang, H.; Li, H.; Chen, H. Research on safety risk transfer in subway shield construction based on text mining and complex network. *Buildings* **2023**, *13*, 2700. [\[CrossRef\]](#)
42. Liu, Y.; Wang, J.; Tang, S.; Zhang, J.; Wan, J. Integrating information entropy and latent Dirichlet allocation models for analysis of safety accidents in the construction industry. *Buildings* **2023**, *13*, 1831. [\[CrossRef\]](#)
43. Tixier, A.J.-P.; Hallowell, M.R.; Rajagopalan, B.; Bowman, D. Application of machine learning to construction injury prediction. *Autom. Constr.* **2016**, *69*, 102–114. [\[CrossRef\]](#)
44. Qiao, J.; Wang, C.; Guan, S.; Shuran, L. Construction-accident narrative classification using shallow and deep learning. *J. Constr. Eng. Manag.* **2022**, *148*, 04022088.
45. Kim, T.H.; Chi, S.H. Accident case retrieval and analyses: Using natural language processing in the construction industry. *J. Constr. Eng. Manag.* **2019**, *145*, 04019004. [\[CrossRef\]](#)
46. Tian, D.; Liu, H.; Chen, S.; Li, M.; Liu, C. Human error analysis for hydraulic engineering: Comprehensive system to reveal accident evolution process with text knowledge. *J. Constr. Eng. Manag.* **2022**, *148*, 04022093. [\[CrossRef\]](#)
47. Kifokers, D.; Xenidis, Y. Application of Linguistic clustering to define sources of risks in technical projects. *ASCE-ASME J. Risk Uncertain. Eng. Syst. Part A Civ. Eng.* **2018**, *4*, 04017031. [\[CrossRef\]](#)
48. Chi, N.W.; Lin, K.Y.; Hsieh, S.H. On effective text classification for supporting job hazard analysis. *ASCE Int. Workshop Comput. Civ. Eng.* **2013**, *2013*, 613–620.
49. D'Orazio, M.; Giuseppe, E.D.; Bernardini, G. Automatic detection of maintenance requests: Comparison of Human Manual Annotation and Sentiment Analysis Techniques. *Autom. Constr.* **2022**, *134*, 104068. [\[CrossRef\]](#)

50. Park, L. KoNLPY: Korean NLP in Python. 2014. Available online: <https://konlpy.org/en/latest/> (accessed on 3 April 2023).
51. Zou, Y.; Kiviniemi, A.; Jones, S.W. Retrieving similar cases for construction project risk management using natural language processing techniques. *Autom. Constr.* **2017**, *80*, 66–76. [[CrossRef](#)]
52. Collins, A.M.; Loftus, E.F. A spreading activation theory of semantic processing. *Psychol. Rev.* **1975**, *82*, 407–428. [[CrossRef](#)]
53. Caldas, C.H.; Soibelman, L. Automating hierarchical document classification for construction management information system. *Autom. Constr.* **2003**, *12*, 395–406. [[CrossRef](#)]
54. Garwood, K.C.; Jones, C.; Clements, N.; Miori, V. Innovations to identifying the effects of clear information visualization: Reducing managers time in data interpretation. *J. Vis. Lit.* **2018**, *37*, 40–50. [[CrossRef](#)]
55. Xiong, Y.; Cho, M.; Boatwright, B. Hashtag activism and message frames among social movement organizations: Semantic network analysis and thematic analysis of Twitter during the #MeToo movement. *Public Relat. Rev.* **2019**, *45*, 10–23.
56. Kang, G.J.; Ewing-Nelson, S.R.; Mackey, L.; Schlittl, J.T.; Marathe, A.; Abbas, K.M.; Swarup, S. Semantic network analysis of vaccine sentiment in online social media. *Physiol. Behav.* **2017**, *35*, 3621–3638. [[CrossRef](#)]
57. Zarei, B.; Sharifi, H.; Chaghoei, Y. Delay causes analysis in complex construction projects: A Semantic Network Analysis approach. *Prod. Plan. Control.* **2018**, *29*, 29–40. [[CrossRef](#)]
58. Fariña García, M.C.; Nicolás De Nicolás, V.L.D.; Blanco, J.L.Y.; Fernández, J.L. Semantic network analysis of sustainable development goals to quantitatively measure their interactions. *Environ. Dev.* **2021**, *37*, 100589. [[CrossRef](#)]
59. Jeon, S.-W.; Kim, J.-Y. An exploration of the knowledge structure in studies on old people physical activities in Journal of Exercise Rehabilitation: By semantic network analysis. *J. Exerc. Rehabil.* **2020**, *16*, 69–77. [[CrossRef](#)] [[PubMed](#)]
60. Okamoto, K.; Chen, W.; Li, X.Y. Ranking of closeness centrality for large-scale networks. In *International Workshop on Frontiers in Algorithmics*; Springer: Berlin/Heidelberg, Germany, 2008; pp. 186–195.
61. National Korean Language Institute. Available online: https://stdict.korean.go.kr/main/main.do#main_logo_id (accessed on 3 April 2023).
62. Goodmanmi. GitHub—Goodmami/wm: A Modern, Interlingual Wordnet Interface for Python. 2023. Available online: <https://github.com/goodmami/wm> (accessed on 10 May 2023).
63. Collins Thesaurus. Available online: <https://www.collinsdictionary.com/dictionary/english-thesaurus> (accessed on 10 May 2023).
64. Condos, J.; Sorrell, W.H.; Donegan, S.L. Blockchain Technology: Opportunities and Risks. Available online: https://sos.vermont.gov/media/253f2tpu/vermontstudycommittee_blockchaintechology_opportunitiesandrisks_finalreport_2016.pdf (accessed on 11 November 2023).
65. Mundt, M.; Majumder, S.; Murali, P.; Panetsos, V.R. Meta-learning convolutional neural architectures for multi-target concrete defect classification with the concrete defect bridge image dataset. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Long Beach, CA, USA, 15–20 June 2019; pp. 11196–11205.
66. Hoang, N.-D. Image processing-based recognition of wall defects using machine learning approaches and steerable filters. *Comput. Intell. Neurosci.* **2018**, *2018*, 7913952. [[CrossRef](#)] [[PubMed](#)]
67. Yassin, A.A.; Hamzeh, F.; Sakka, F.A. Agent based modeling to optimize workflow of robotic steel and concrete 3D printers. *Autom. Constr.* **2020**, *110*, 103040. [[CrossRef](#)]
68. Garza, J.M.; Pishdad-Bozorgi, P. Workflow process model for Flash Track Projects. *J. Constr. Eng. Manag.* **2018**, *144*, 06018001. [[CrossRef](#)]
69. Azeez, M.; Gambatese, J.; Hernandez, S. What do construction workers really want? A study about representation, importance, and perception of US construction occupational rewards. *J. Constr. Eng. Manag.* **2019**, *145*, 04019040. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.