Article

Construction Disputes and Associated Contractual Knowledge Discovery Using Unstructured Text-Heavy Data: Legal Cases in the United Kingdom

JeeHee Lee [1] , **Youngjib Ham** [2] **and June-Seong Yi** [3],*

1   Department of Civil and Environmental Engineering and Construction, University of Nevada Las Vegas, Las Vegas, NV 89154, USA; jeehee.lee@unlv.edu
2   Department of Construction Science, Texas A&M University, College Station, TX 77843, USA; yham@tamu.edu
3   Department of Architectural and Urban Systems Engineering, Ewha Womans University, Seoul 03760, Korea
*   Correspondence: jsyi@ewha.ac.kr

Abstract: Construction disputes are one of the main challenges to successful construction projects. Most construction parties experience claims—and even worse, disputes—which are costly and time-consuming to resolve. Lessons learned from past failure cases can help reduce potential future risk factors that likely lead to disputes. In particular, case law, which has been accumulated from the past, is valuable information, providing useful insights to prepare for future disputes. However, few efforts have been made to discover legal knowledge using a large scale of case laws in the construction field. The aim of this paper is to enhance understanding of the multifaceted legal issues surrounding construction adjudication using large amounts of accumulated construction legal cases. This goal is achieved by exploring dispute-related contract terms and conditions that affect judicial decisions based on their verdicts. This study builds on text mining methods to examine what type of contract conditions are frequently referenced in the final decision of each dispute. Various text mining techniques are leveraged for knowledge discovery (i.e., analyzing frequent terms, discovering pairwise correlations, and identifying potential topics) in text-heavy data. The findings show that (1) similar patterns of disputes have occurred repeatedly in construction-related legal cases and (2) the discovered dispute topics indicate that mutually agreed upon contract terms and conditions are import in dispute resolution.

Keywords: construction dispute; contract management; knowledge discovery; text mining

1. Introduction

The construction industry has witnessed an increase in the number of litigation cases arising from disputes over the last few years [1–4]. According to the report by the National Bureau of Standards [5], 30% of construction firms were involved in at least one dispute over a 12-month period. If a construction dispute cannot be resolved through negotiation and compromise between the parties, it becomes necessary for an impartial third party to resolve the disagreement [6,7]. Litigation and arbitration are common dispute resolution procedures, but they are costly and time-consuming [8]. Recent reports including [9] show that an average of $42.4 million has been spent on each dispute resolution in the global construction market. This demonstrates that construction disputes are one of the main factors that prevent construction projects from being completed successfully [1]. To improve sustainability in construction, undesirable economic and social impacts of construction disputes should be avoided, taking into account uncertainty.

Many studies have found that one of the notable reasons for construction disputes is contractual issues that could have been avoided [10–13]. In a recent report by Arcadis [9], poor contract administration was identified as one of the main causes of construction

disputes. Since contractual facts stated in contracts are binding agreements between the parties and are important pieces of evidence for legal judgments [14], clearly expressed contract conditions are essential to prevent conflicts of a contractual interpretation. Moreover, as the conditions of construction contracts have become increasingly complex and sophisticated [10], some parties often have hidden provisions in contracts that can be beneficial to them [3]. Hughes [15] pointed out that some owners tend to maliciously modify or even remove the terms of standard conditions, thereby shifting the risks to contractors. In practice, disputes caused by materially adverse contracts are often found in litigation [11]. Therefore, avoiding construction disputes requires understanding of the contractual terms and relationships that exist on a construction project [12].

Given the nature of the experience-oriented construction industry, knowledge and judgments of past experiences have been considered key factors in preventing and resolving disputes and associated problems that may occur (e.g., delays, cost overruns) [16]. In particular, accumulated case law is valuable information that provides useful lessons for future disputes. Case law, which plays a critical role in legal reasoning and decision making, is the corpus of decisions on cases that judges have made [17]. As every legal judgement is made on the basis of contractual evidence [18], contract conditions that are cited in the legal judgement can be useful to understand potential risk factors. In other words, frequently appearing contract conditions in legal cases can be critical contract terms that should be carefully reviewed at the contract level because courts seem to rely on contract interpretation whenever disputes result in litigation [18,19]. Therefore, by examining construction legal cases focusing on contract terms and provisions, there is the potential to obtain meaningful contractual knowledge that can provide a better understanding of avoiding uncertainties, reducing future disputes.

Prior works on construction legal cases including [18,20–22] have mainly focused on manual content analysis of construction cases and review of specific reports that identified the disputes. Since large-scale manual analysis of text and content is time-consuming, error prone, and subject to data bias, little research exists on exploring large amounts of case laws. Moreover, the case base is comprised of a large number of cases and grows every year, thus the legal professional faces difficulties in retrieving and interpreting information from the case base [17]. Advancements in text analysis technologies have made possible the discovery of previously unknown but potentially useful knowledge in large amounts of unstructured text data [23]. For example, text mining-based reviews of research articles have been performed in the construction domain [23,24] to identify research trends of specific issues from unstructured or semi-structured textual data. A major benefit of using text mining lies in the possibilities of discovering document topics and detailed properties and relationships from a massive volume of text data. Given the nature of construction projects, which have different distinctions from project to project and cannot be replicated so that the same tasks and resources provide the same results, it is noted that as more cases are analyzed, more diverse and meaningful knowledge can be obtained.

This study aims to provide a better understanding of the multifaceted legal issues surrounding construction disputes through a bottom-up investigation starting from the legal cases. To achieve the goal, we explore knowledge about dispute-related contract terms and commonly occurring patterns that affect judicial decisions based on their verdicts by automatically identifying legal arguments, properties, and relationships in accumulated legal cases from a long period of time. This study builds on text analytics to identify the types of disputes in large amounts of legal cases without examining hundreds of cases manually. Given that construction practitioners may need to understand the dynamics of complex construction law in order to deal with construction conflicts and disputes, text analytics and natural language processing (NLP) can play a significant role in this field. Therefore, such text-based automatic analysis will provide a powerful and scalable approach and thus expand accessibility in legal information, particularly to those with little knowledge of contract management and legal information. This study aims to facilitate

construction practitioners in actively engaging in understanding construction legal issues using text analytics.

The collected legal dispute cases were categorized by contractual topics, based on the text topic modeling method, latent Dirichlet allocation (LDA). The results of this study are expected to contribute to expanding a body of knowledge of sustainability in construction by providing (1) interpretable and visualized outcomes that can help to understand construction dispute issues while minimizing human effort; (2) accessibility of legal information and efficient use of such knowledge for contract management; and (3) advanced understanding of contractual knowledge that helps to avoid uncertainties and thus prevent future disputes proactively. The paper begins by providing a review of previous studies on construction disputes and knowledge discovery methods and discusses research gaps to be filled in this study. Then, the proposed framework is explained in detail. Finally, construction disputes and related contractual issues are examined based on the proposed framework and the experimental results are discussed.

2. Literature Review

Construction stakeholders are frequently faced with a variety of conflicts and disputes during a project. There are differences between conflicts and disputes; conflicts can be managed among construction parties, but disputes, resulting from conflicts, are associated with distinct justiciable issues and require third party interventions such as courts and arbiters [20,25].

Given the nature of the experience-driven construction project, learning from previous experiences and failures has been considered valuable; however, once a project is completed, the associated experiences are likely inaccessible, and the learning opportunity to the wider community wasted, only remaining as tacit or experiential knowledge [26]. Previous studies have put forth models and examined factors to enhance the understanding of what causes construction disputes. For example, Semple et al. [12] pointed out that if one party feels that the contractual obligations or expectations have not been met, and if there is disagreement between the parties, construction disputes occur. They presented the results of a pilot study of 24 construction claims in order to identify common causes of claims and problem areas within the construction process. In addition, Mitkus and Mitkus [27] investigated the causes of conflicts arising between clients and contractors and found that a contract allowing room for different interpretations by the parties is one of the main causes of conflicts in construction projects. Abdul-Malak and Senan [28] investigated factors and trends that have been governing the implementation of adjudication in construction practice. However, the authors explored a total of 28,000 references manually, thus reducing the opportunity for knowledge reuse. Information generated in construction projects is usually unstructured and text-heavy, which is disadvantageous in terms of data utilization efficiency compared to structured numeric data. Despite the importance of text data in construction projects, transferring tacit and experiential knowledge from construction documents to practitioners has been limited, thereby reducing knowledge reuse and sharing.

With the advancement of unstructured text data analytics, recent studies have attempted to analyze text information from vast amounts of documents using text mining techniques. As part of these efforts, Al Qady and Kandil [29] presented a knowledge discovery that extracts contractual knowledge in contract documents. The proposed methodology facilitates quick access and efficient use of such knowledge for project management and contract administration. In construction dispute management, historical dispute cases have been retrieved in order to easily obtain relevant information from unstructured text documents. For example, Liu et al. [16] proposed a dispute settlement model for international construction projects based on case-based reasoning (CBR), mining successful experiences in historical dispute settlement cases. Fan and Li [30] applied text mining to find similar construction accidents occurring in the past for resolving disputes through alternative dispute resolution. Jallan et al. [31] also built on natural language processing

and text mining to identify frequently appearing patterns and themes associated with construction-defect litigation. Previous studies have indicated that discovering useful knowledge hidden in construction documents or historical cases is valuable for achieving a successful project. While several studies including Abdul-Malak and Senan [28] have shown the feasibility of knowledge discovery using historical dispute data (i.e., legal cases), a detailed analysis of dispute-related contract clauses that govern litigation outcomes has not been fully investigated with large corpora. Being aware of potential contractual risks ahead will help construction practitioners when drafting contracts, but the large scale of legal data has made it difficult to explore commonly occurring patterns and themes that affect judicial decisions from the legal cases. Given that having solutions for construction disputes and litigation begins with an understanding of critical contract terms and clauses, analysis of a large number of legal cases focusing on contract terms and provisions reveals meaningful contractual tacit knowledge to prevent future disputes.

3. Research Method: Construction Disputes and Related Contractual Knowledge Discovery

This study presents a data-driven framework to identify the causes of construction legal disputes, specifically focusing on contractual issues, with the aim of providing knowledge-inspired dispute causes related to contract issues using unstructured and text-heavy data. Since a case study is used to explain, describe, or explore events or phenomena in the everyday contexts in which they occur [32], this case study explores multifaceted and complex legal issues in construction using UK legal cases. According to Arcadis [9], the UK continues to experience most disputes being resolved after they have crystallized, rather than parties seeking to avoid or mitigate a potential dispute situation. Thus, investigating the legal cases of UK construction industry as a case study will provide a better understanding and possible solutions to dealing with complex legal issues.

Figure 1 illustrates the research process used to address the research goal and R programming language with packages such as *Tidytext* and *LDAvis* has been used to implement the entire text data processing.
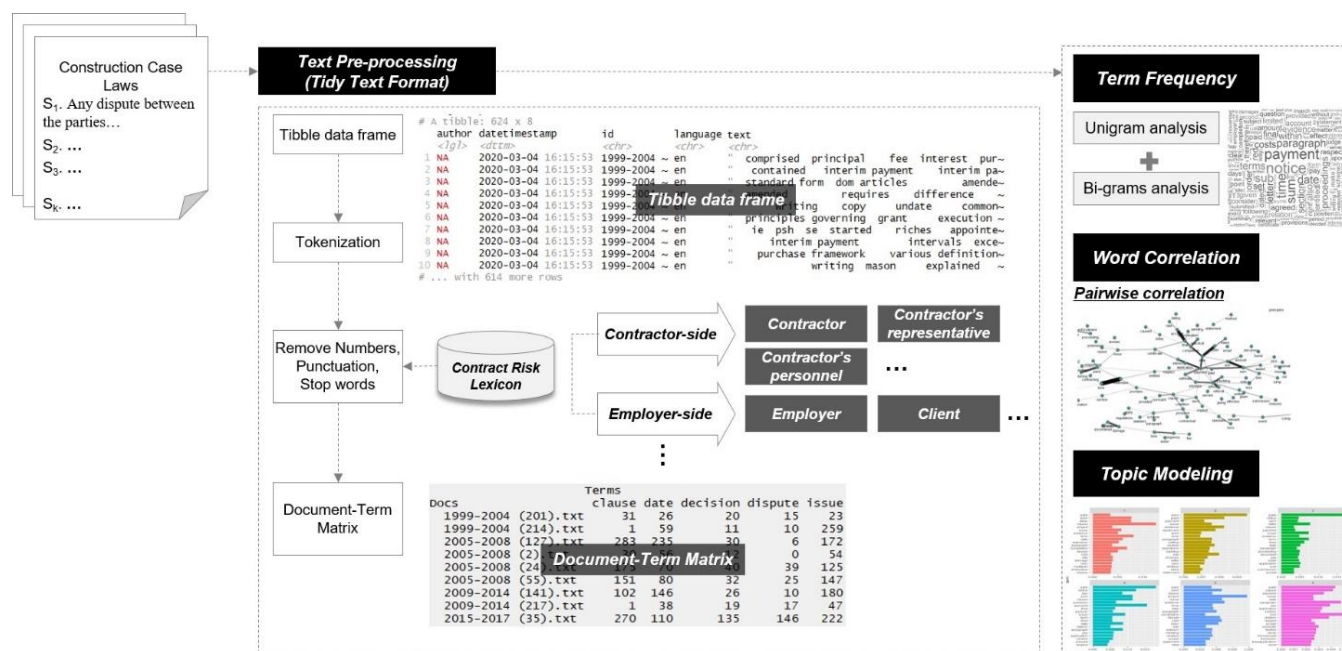

Figure 1. Research process.

The first step is pre-processing of collected legal cases. This study used construction adjudication cases as the main data source for text analysis. A total of 624 dispute cases were collected between 1999 to 2019 from the British and Irish Legal Information Institute

(BAILII) and an adjudication nomination body of the UK and all legal cases associated with construction contracts were crawled from the webpage [33]. Cases related to other types of contracts (e.g., facility management contracts, maintenance contracts) were excluded from the analysis. Most of the contract forms in the collected data were under Joint Contracts Tribunals (JCT), one of the most common forms of construction contracts used in the UK. On average, the legal cases collected were 10–11 pages in PDF format, and the documents usually consisted of four parts: (1) an introductory section that provides a preamble and a dispute background, (2) the relevant provisions of the contract, (3) issues that arise among parties containing the elements that describe the disputes, and (4) decisions and conclusions of the disputes.

The collected legal dispute cases in the unstructured text format were converted to a tidy data frame format, which is a table with one token (i.e., basic meaningful unit of text) per row. Using tidy data is a powerful way to make handling data more effective, and this is no less true when it comes to dealing with text [34]. A set of text documents is first segmented by sentence unit, and tokenization is performed, which is the process of splitting text into tokens. The initial text format is a linear sequence of words or phrases, which can be segmented into linguistic units by tokenization, making it easier to analyze. Several stop-words, including numbers, punctuation marks, and meaningless symbols are removed through pre-processing. Construction legal cases are a special kind of textual data with many legal terms; thus, some of the recurring legal terms (e.g., plaintiff, jurisdiction, adjudication) that are less important to the inference of dispute causes need to be cleaned to obtain robust results. Text pre-processing is performed to remove unnecessary information from text documents as well as to structure the data format so that the computer can easily understand it.

To better understand semantic relationships and hierarchy of words in legal cases, we built upon the contract domain lexicon [3,35] in the text pre-processing step. The lexicon was developed to identify vocabularies used in the construction contract domain, and it has seven sub-classes (environment, resource, actor, payment, action, process, and right and responsibility) that are used to categorize each term as relating to the contractual use. The lexicon allows for text pre-processing effectively by providing a domain-dependent lexical dictionary. For example, 'change order' and 'variation' can generally be used interchangeably when a mutually agreed scope of work in a construction contract is altered. The developed lexicon puts these two terms into a single class so that they can be within similar topics. Once pre-processing is complete, the collected data are ready for further text analysis. To discover dispute-related knowledge from the text data, frequently used terms and word correlations are discovered. The results of analyzing frequently used terms can provide a general understanding of issues that often arise in legal cases. In order to better understand potential contractual issues (i.e., topics) underlying legal cases, topic modeling is performed based on a document-term matrix whose rows represent the documents in the collection and columns represent terms.

3.1. Frequently Occurring Dispute Issues

As the first step of the text analysis of construction legal cases, we investigated the overall issues in construction disputes by analyzing frequently used terms. Examining frequently used terms in a given text dataset provides a general understanding of frequent issues listed in legal documents. Figure 2 shows the most frequently used words, which occurred more than 4000 times in a total of 624 documents from 1999 to 2019. The most frequently occurring word was 'clause', which occurred 11,803 times in the legal cases, followed by 'payment', and 'issue'. Despite only a simple frequency count, the results enable insightful interpretation. For example, a frequently used term such as 'notice' provides a clue to infer how important the obligation of notice is in communication among parties. It shows that the presence or absence of notice in the process of a construction project can be a significant part of the final legal decision of a construction dispute.
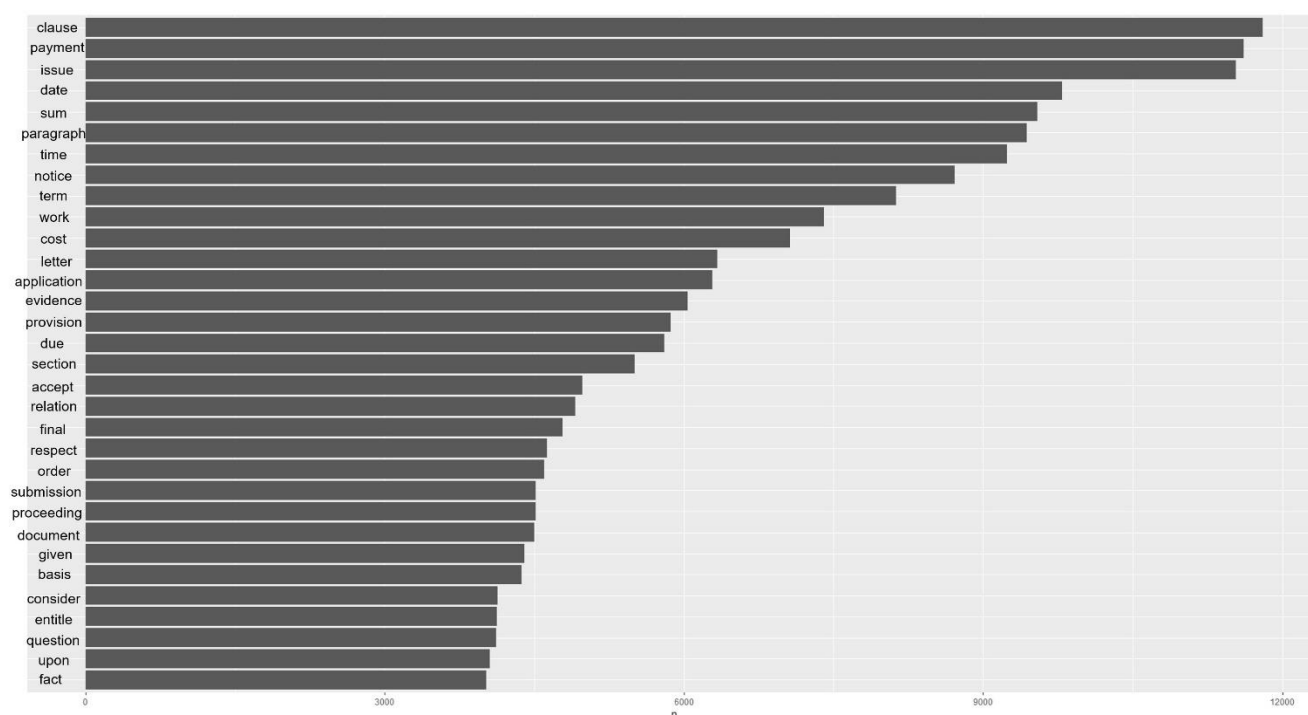
Figure 2. Frequent terms in the construction legal cases.

Although a single term gives straightforward information, additional supplemental words that explain the context in text data need to be studied. The keyword 'payment' is a good example that shows why additional analysis is needed; for example, 'payment due', 'withhold payment', and 'interim payment' each have different meanings, which depend on the combination of 'payment' with other words. To address this, we combined adjacent words together for tokenization, based on which more interpretable results can be obtained [23], in order to discover the relationships among adjacent terms in the given documents.

Table 1 shows the top 10 frequent terms by unigrams and bigrams, respectively. As shown in Table 1, the most frequent unigram 'clause' does not appear in the list of the top 10 frequent bigrams, because the adjacent words used with 'clause' are mostly numbers (e.g., clause 1.1, clause 1.2) that were removed in the text pre-processing. 'Extension time', which is the most frequently used bigram in the legal cases (a total of 2383 times in the cases), is a common issue in construction projects.

Table 1. Top 10 frequent terms (unigrams and bigrams).

Unigram	# of Occurrence	Bigram	# of Occurrence
Clause	11,803	Extension time	2383
Payment	11,610	Sum due	1369
Issue	11,538	Final account	1334
Date	9793	Letter date	1314
Sum	9541	Interim payment	1052
Paragraph	9440	Liquidated damages	1014
Time	9239	Date payment	906
Notice	8714	Practical completion	902
Term	8129	Withholding notice	848
Work	7407	Loss expense	837

#: number of . . .

Almost every construction project faces delays resulting in extension of time (EOT), and the terms of the contract bind which of the two contract parties absorbs the responsibility for the delay, in which there is room for interpretation of the contract [36]. Moreover, 'liquidated damages' caused by the extension of time was listed as a frequently occurring

bigram in the legal cases, indicating that a construction delay-related issue is one of the biggest dispute issues in construction projects. Another frequent bigram, 'final account', denotes a contractual concept, which is the conclusion of the contract sum (i.e., including all necessary adjustments) and signifies the agreed amount that the employer will pay the contractor on completion of the work written in the contract [37]. Most standard contracts for construction projects include clauses dealing with final accounts in order to impose a consensus on the possible conflicts arising from quality of work, decisions on extension of time, and loss or expense. Therefore, the results of term frequency show how many legal cases there are in relation to the final account issues. However, reviewing frequent terms needs to be interpreted carefully because the number of occurrences of the bigram 'extension time (EOT)' does not necessarily mean that there are 2383 disputes related to EOT. Rather, it can be interpreted that EOT was discussed by the majority of verdicts in legal cases; therefore, further analysis was conducted in the following section.

Discovering terms that are not adjacent but concurrent can also help to give useful insights because several terms, which are key clues to understand the meaning of a sentence, can occur concurrently, but not adjacently in the same sentence [38]. The pairs of words that occur together most often were investigated through pairwise correlation focusing on the phi coefficient, a common measure for binary correlation considering how much more likely it is that either both words A and B appear, or that neither do, than that one appears without the other [34].

Figure 3 illustrates networks of co-occurring words in order to better understand relationships between words. As shown in Figure 3, several words are connected around the term 'payment'. In particular, weak clusters are connected with each other in the networks; as an example, contractual notice-related terms (e.g., notice, written, effective) are clustered with each other and are linked to another cluster such as contract clause-related terms (e.g., clause, condition, provision, term). This shows that the contractual notice-related terms and the contract clause-related terms frequently occurred together in the legal cases. Given that several notice obligations of contract parties are stated under a JCT building contract, this verifies that one of the parties' obligations, the obligation of notice, should be obeyed in accordance with the contract statement, which is an important factor when a legal decision is made. In addition, the link between the notice-related terms and payment-related terms (e.g., payment, pay, less) shows that the payment notices for construction contracts (e.g., pay less notice) frequently occur in the case of legal disputes.
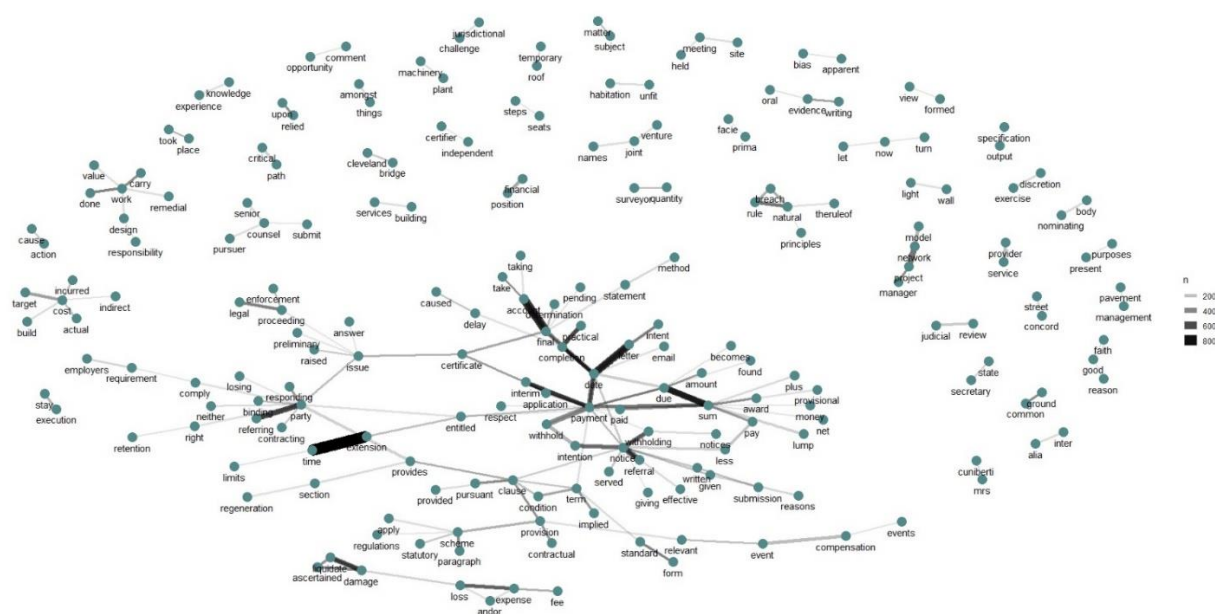

Figure 3. Pairwise correlation of words (*number of occurrences of pairwise words >70*).

Figure 4 shows another perspective of the observed word correlations. Positive correlations are displayed in blue and negative correlations in red circles. Color intensity and the size of the circle are proportional to the correlation coefficients [39]. Since the matrix is mirrored around its diagonal, correlation values of any two attributes of both sides are the same based on the diagonal. For example, Figure 4 indicates that 'submission' has a strong positive correlation with 'clear' because it has a large blue circle at the intersection between 'submission' and 'clear'. If we look at the correlation in terms of 'clear', strong positive correlation with 'submission' can be observed as well. The strong correlation between 'submission' and 'clear' means that clarity of information should be met when the contractor submits construction documents.
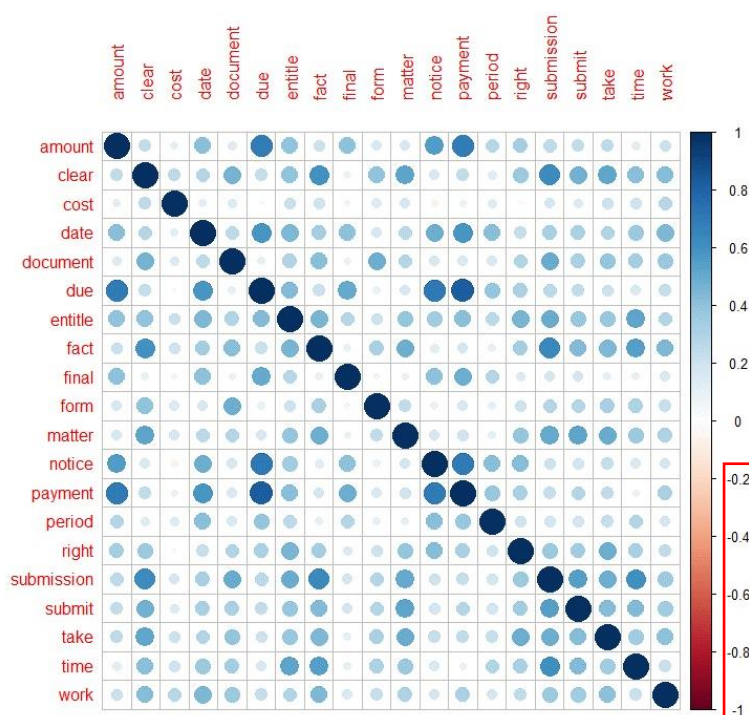

Figure 4. Word correlation in legal dispute cases.

3.2. Construction Dispute Trends by Period

There have been findings that the economic circumstances surrounding a construction project can affect the number of construction claims and disputes [40,41]. Therefore, analyzing the construction disputes in the context of the construction economic situation is needed for the trend analysis. Considering that disputes may vary under different circumstances, the collected legal cases were grouped based on the construction industry trends in the UK from 1999 to 2019. A total of 624 legal cases were grouped into four different periods based on the growth of the construction market during 1999–2019 [42] as follows:

- First group (1999 to 2004): the period during which the size of the construction market grew,
- Second group (2005 to 2009): the period during which the size of the construction market reduced,
- Third group (2010 to 2014): the period during which the size of construction market fluctuated, and
- Fourth group (2015 to 2019): the period during which the size of the construction market was stable.

Based on the results of Section 3.1, it is noted that contract conditions and clauses are frequently referenced in the legal cases and are relevant to legal decisions. This demon-

strates that mutually agreed contract terms and conditions play an important role in the process of legal dispute resolution. Understanding which contract conditions appear often in the legal dispute documents, especially in the final legal decisions, has the effect of being able to help prevent future disputes. Therefore, in order to focus on the contractual issues that affect the legal decisions of disputes, we extracted only the "relevant provision of the contract" and "decisions and conclusions of the disputes" from each dispute case. Using the extracted data, we investigated dispute trends by period and identified potential topics in text data.

Figure 5 depicts what aspects of disputes have been addressed in four different periods. The results show that similar issues occurred repeatedly in construction legal cases throughout the entire period, showing that there have been similar historical experiences that can be used to solve the present issues. As shown in Figure 5, delay-related disputes such as 'extension time' and 'liquidated damages' are still main issues throughout the entire period. In addition, 'practical completion', which is also referred to as 'substantial completion', appears equally in every group. A construction project is considered to be practically complete when there are no outstanding defects (except for minor items or snagging) and the building or infrastructure can be put to its intended use [43]. Practical completion is an important milestone because it means that the contractor is entitled to the release of any retainage, excluding deductions for uncompleted works, and the contractor is no longer liable for liquidated damages beyond practical completion [44]. Given that almost every legal case collected in this study is under the JCT contract that has no definition of practical completion, it can be understood why the term 'practical completion' has been referenced often as one of the causes of disputes in the collected legal cases. In addition, the fact that 'commissioning program' appeared in the 3rd period (2010 to 2014) implies that commissioning-related issues have increased in construction legal cases as the practice of construction commissioning has matured recently. The period from 2015 to 2019 seems to have slight tendencies compared to other periods, since it has several strong issues associated with 'interim payment' and 'notice'. This shows that disputes caused by failing to issue an adequate and timely interim certificate (i.e., a payment notice) have occurred more frequently recently. Additionally, several common words (e.g., 'project network', 'network model'), which are not shown in Figure 5 but still appear often in the 4th group, indicate that there have been some disputes regarding a project's scheduling model.
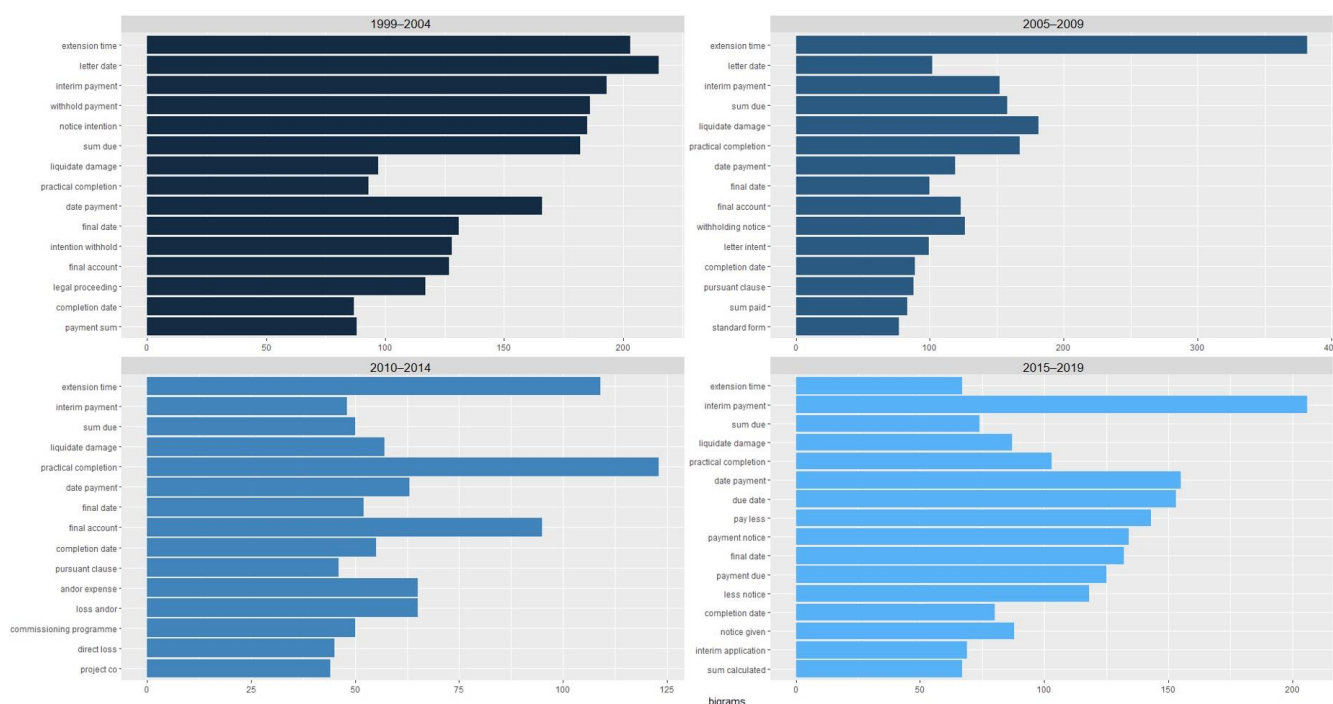

Figure 5. Frequently used terms by period.

Figure 6 illustrates how the major issues (i.e., frequent terms) change over time. As shown in Figure 6, several major issues, such as 'extension time' and 'practical completion', appear steadily over time; however, payment-related issues (e.g., 'interim payment', 'payment due') became more likely to be referenced in the legal decisions. Payment-related contract conditions and clauses became more likely to be cited in the course of legal decisions, which means that payment-related contract conditions should be carefully examined during the contracting phase.
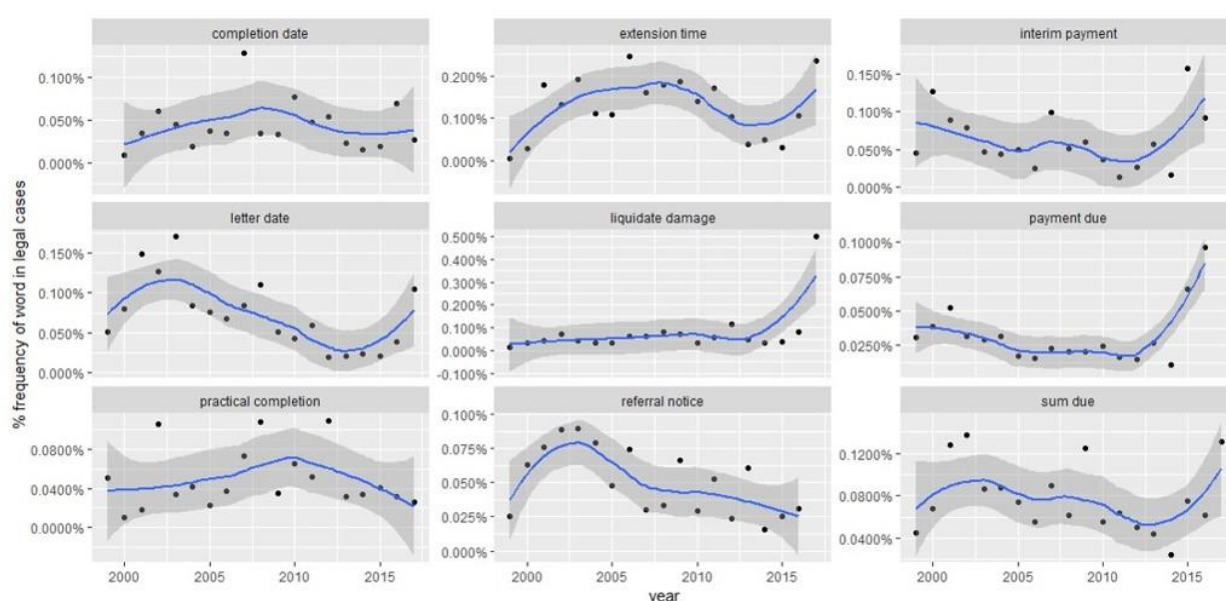

Figure 6. Changes in word frequency of construction legal cases over time.

3.3. Topics Discovered

To discover the underlying themes or topics in the collected legal cases, text topic modeling, which organizes a collection of documents based on the discovered topics, was conducted. Unlike the term frequency analysis, text topic modeling allows text in a document to be classified to a particular topic so that we can better understand hidden themes from documents. To facilitate text topic modeling, we built on LDA, which is a generative probabilistic model of a corpus and is based on the idea that documents are mixtures of topics, where each topic is represented by a probability distribution over words [45]. Unlike naïve Bayes classifiers, which assume that each document is on a single topic, LDA lets each word be on a different topic. In the process of LDA, there is an assumption that the K latent topics are hidden in the given collection of text documents D, and each topic is represented as a multinomial distribution over all the M words. LDA assumes the following generative process for a corpus D consisting of N documents [45–48]:

1. For every document d in the whole document-corpus D, randomly choose a distribution over topics; i.e., choose $\theta_n \sim$ Dirichlet (α), where $n \in \{1, \ldots, N\}$ with a symmetric parameter α which typically is sparse ($\alpha < 1$).

2. For each word $w_{n,m}$ in the document d:

    (a) Randomly choose a topic from the distribution over topics: i.e., choose a topic $z_{n,m} \sim Multinominal(\theta_n)$

    (b) Randomly choose a word from the corresponding topic (distribution over the vocabulary):
    i.e., find the corresponding topic distribution $\varphi_{z_{n,m}} \sim Dirichlet(\beta)$ and a symmetric parameter β typically is sparse; and

    (c) Sample a word $w_{n,m} \sim Multinominal\left(\varphi_{z_{n,m}}\right)$.

By repeating the above procedures N times, document d can be generated word by word and the whole corpus can be generated. Taking the product of the marginal probabilities of each document, the probability of a whole corpus is estimated:

$$P(D|\alpha, \beta) = \prod_{n=1}^{N} \int P(\theta_n|\alpha)F(\theta, \varphi)d\theta_n, \tag{1}$$

$$\text{where, } F(\theta, \varphi) = \prod_{m=1}^{M} \sum_{z_{n,m}} P(z_{n,m}|\theta_n)P(w_{n,m}|\varphi_{z_{n,m}})$$

The topic-word distributions φ and document-topic distributions θ in Equation (1) could be learned through the Gibbs sampling and expectation-maximization (EM) algorithm via maximizing the probability $P(D|\alpha, \beta)$ [46]. Previous studies on text topic modeling have shown that if the overall length of the document is too long, it is likely to be problematic for topic models since it includes an arbitrary mixture of too many topics in a document and makes it difficult to identify meaningful topics [49,50]. Therefore, we split the legal cases that have four parts (i.e., introductory section, relevant provisions of the contract, issues arising among parties, and decisions and conclusions of the disputes) into thematically coherent parts, and only the part of the relevant provisions of the contract and the decisions of the disputes were used for the topic modeling to focus on final legal decisions and corresponding contract evidence. For that purpose, additional stop words referring to the contract clauses (e.g., condition, clause, provision) were removed.

Figure 7 indicates the discovered topics through text topic modeling. Topic modeling was performed using a unigram document-term matrix whose rows represent the documents in the collection and columns represent unigram terms [51]. The reason why the unigrams, not the bigrams, were applied for the topic modeling analysis is that most bigram words, except for some frequent words, were too sparse to get meaningful results in the topic modeling analysis. A total of seven topics were finally obtained using the harmonic mean method to determine the optimal number of topics. The scalar value of the Dirichlet hyper-parameter for topic proportions, alpha, was set at 0.02 and the number of iterations was 2000 in our study. Several terms such as 'payment' were assigned more than one group based on their pairwise correlations. For example, 'payment' showed a high correlation coefficient with the terms 'amount' and 'total', which are part of Topic 7. At the same time, 'payment' had high correlation with 'notice' and 'date', which are part of Topic 6. Each topic has been labeled as shown in Table 2 and the seven key aspects of construction disputes are discussed in detail in the following subsection.

3.3.1. Topic 1: Letter of Intent (LOI)

Topic 1 has been labeled as a 'letter of intent (LOI)' based on the most salient terms in this group. Topic 1 consists of terms such as 'letter', 'evidence', 'intent', and 'signed'. It indicates that the legal dispute cases are associated with issues of the letter of intent. A letter of intent is a document expressing the employer's intention to enter into a formal written contract for works described in the letter, and asking the contractor to begin those works before the formal contract is executed [52]. The term 'letter of intent' itself has no legal meaning, but the legal effect of a letter of intent may differ depending on the particular meaning and purpose defined in each project. Therefore, disputes can arise if contractors are engaged in projects and on site carrying out work in advance of finalizing the formal contract. Topic 1 shows several past disputes related to letters of intent in construction projects, and it demonstrates that contract parties should be careful to make sure that a letter of intent was clearly written and defined the scope and purpose of whether or not it imposed a binding obligation on the parties.
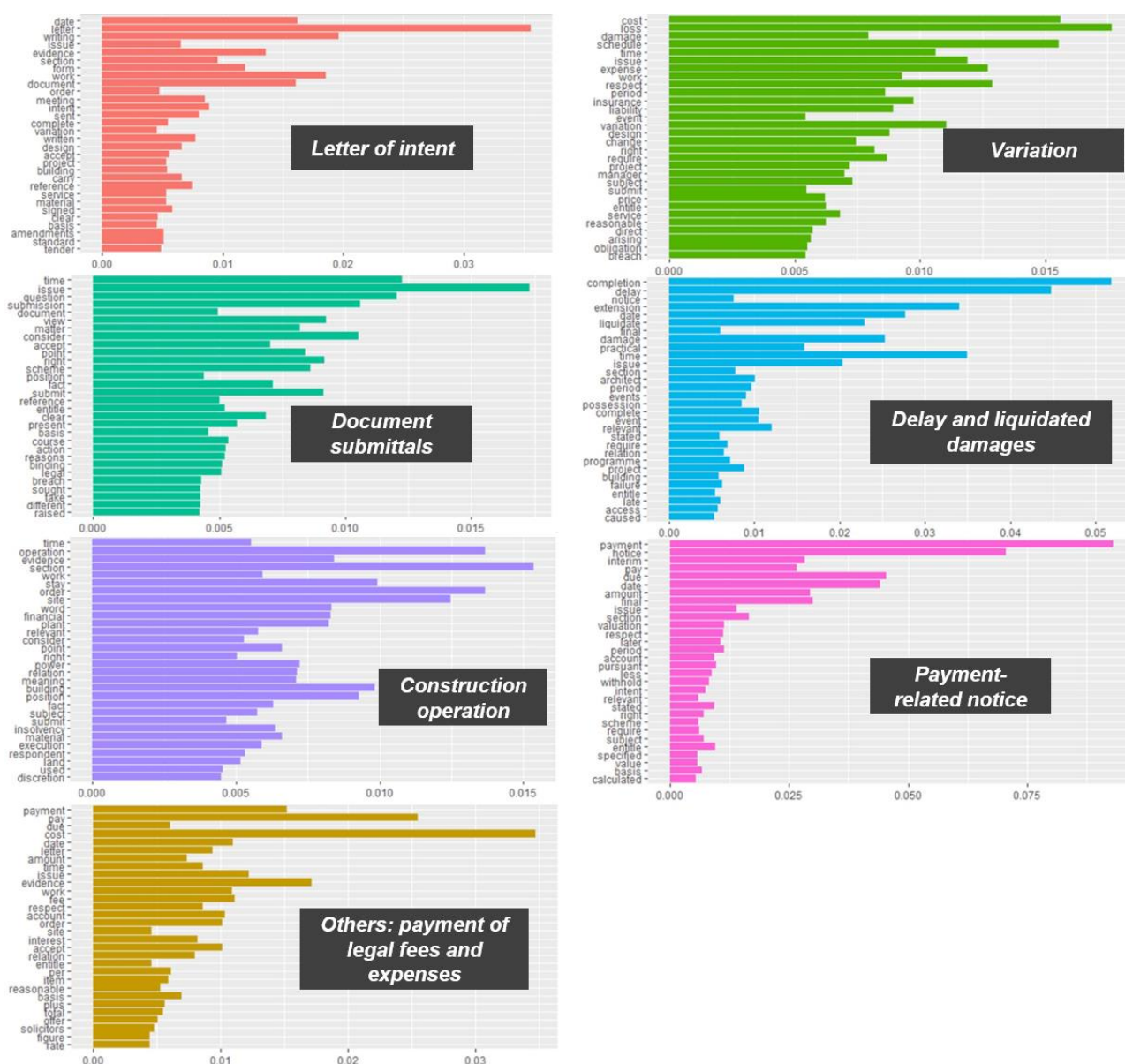
Figure 7. Results of topic modeling analysis.

Table 2. Results of the topic modeling.

Group	Salient Terms
Topic 1 Letter of intent (LOI)	letter, writing, date, work, intent, document, signed, carry
Topic 2 Variation	loss, cost, schedule, variation, work, change, expense, insurance, liability
Topic 3 Document submittals	issue, time, submission, submit, scheme, document, view, accept, time
Topic 4 Delay and liquidated damages	completion, delay, extension, practical, liquidate, damage, possession, access, building
Topic 5 Construction operation	section, operation, order, site, work, stay, meaning, building land
Topic 6 payment-related notice	payment, notice, due, date, issue, valuation, pay, less, interim
Topic 7 Others: Payment of legal fees and expenses	cost, pay, evidence, payment, amount, fee, reasonable, total, solicitor

3.3.2. Topic 2: Variation

As shown in Figure 7, the most frequent terms of this group are 'loss', 'cost', 'expense', 'variation', 'design', and 'change'. These terms indicate that one of the main issues in the collected legal cases is linked to variation issues and corresponding loss and expense. A variation is a deviation from an agreed upon, well-defined original scope of works in a construction contract and variations are common in all types of construction projects [13]. Variations or change orders are likely to incur construction rework or extra work and they cause issues associated with cost and time, which can lead to conflicts and disputes between parties. Under a JCT standard building contract, the contractor must carry out instructions requiring variations unless it makes a reasonable objection, and valuation of variations must be carried out in accordance with rules in the JCT standard building contract [53]. However, several disputes were discovered in the collected cases when the parties were not in mutual agreement about the valuation of variations carried out by contractors. As an example of the classified with this topic, some conflicts show that contractors have no right to claim for the extra payment for extra work unless such work is stated in the contract clause.

3.3.3. Topic 3: Document Submittals

As shown in Figure 7, there are a few words that help to understand the theme of the group in Topic 3. Frequently occurring terms such as 'submission', 'time', 'document', and 'submit' suggest that this group is related to the 'document submittals' that are one of the contractor's obligations. Document submittals in construction projects should comply with contractual procedures written in contracts. The major issues in which disputes arise among the parties are likely to be associated with time and clarity of documents, in terms of document submittals. If contractors fail to submit construction documents and information in a timely manner, it could violate the terms of the contract and cause problems. In addition, any submissions that contain errors and incomplete information can be rejected by the employer's party; worse, they can be used to demonstrate the contractor's poor performance under a construction dispute. These types of dispute cases have been discovered in this analysis, demonstrating that accurate and timely submission of construction documents can prevent related unnecessary construction disputes.

3.3.4. Topic 4: Delay and Liquidated Damages

Topic 4 has been labeled as 'delay and liquidated damages'. It is common for construction projects to face delays that can be characterized as excusable, non-excusable, or compensable. Therefore, conflicts arising over who is responsible for the delay and whether the contractor is entitled to an extension of time can be considered major issues of construction projects. As shown in Figure 7, terms like 'completion', 'delay', 'time', 'extension', 'liquidate', and 'damage' show that legal cases related to construction delays and liquidated damages are one of the main topics in the given documents. In addition, 'practical completion' seems to be one of the most relevant terms in Topic 4. Few disputes have arisen regarding the criteria for judging whether the contractor has reached the 'practical completion', which is an important and appropriate benchmark for the contractor's completion of obligation.

Moreover, terms like 'possession' and 'access' indicate that 'possession of the site' is an issue in the legal cases of Topic 4. The JCT standard building contract requires the employer to give the contractor possession of the site of the work on the date stated in the contract. If the employer fails to give possession on the stated date, it is a serious breach of contract [53]. Some legal cases have been found to be related to delays caused by the employer's failure to give possession on a specified date. This issue demonstrates that if certain contract conditions intentionally exclude the contract terms associated with the contractor's possession of the site, the contractors cannot be assured in case of delays caused by the employer's failure to give possession.

3.3.5. Topic 5: Construction Operation

Topic 5 includes terms such as 'operation', 'order', and 'site' that invoke corresponding topics. However, it is not easy to infer what a common theme is with these words alone. We thus further investigated associated words of some frequent terms to find additional clues. According to the word association results, the term 'operation' has strong connections with 'mean', 'definition', 'housing', and 'act'. This shows that the definition of construction operation by the Housing Grants, Construction, and Regeneration Act 1996 (HGCRA, also known as the Construction Act) [54] has been referenced many times in the given legal cases. The Act defined several activities (e.g., construction alteration, repair, maintenance, etc.) as construction operations; however, some activities (e.g., drilling for oil or natural gas, extraction of minerals, etc.) are expressly excluded. There are thus some disputes where a contract relates partly to construction operations and partly to being excluded in the construction projects. Issues remain if the dispute relates partly to matters that are not construction operations and there is no express adjudication clause [55]. Therefore, given that terms indicating the scope of construction operations frequently appear in this group, we can learn that the issue of 'construction operation', particularly how to resolve disputes about excluded activities from the construction operations, should be clearly defined in advance to prevent related disputes.

3.3.6. Topic 6: Payment-Related Notice

Topic 6 has been labeled as 'payment-related notice' based on terms like 'payment', 'notice', 'pay', 'less', 'interim', and 'due'. If a construction contract has been made under a JCT contract and the employer fails to issue an adequate and timely interim certificate (i.e., a payment notice), the employer may be in an unnecessary dispute. Furthermore, where the employer's party has failed to serve a valid pay less notice, it results in the employer being obliged to pay the entire amount applied for as per the contractor's payment application (which will be deemed as default payment notice) [56]. In the collected legal dispute cases, there are a few cases in which the pay less notice was late and therefore invalid. The following contract condition is an example of contact terms and a condition that are frequently cited in legal decisions:

> "*If the Employer intends to pay less than the sum stated as due from him in the Payment Notice or Interim Application, as the case may be, he shall not later than five days before the final date for payment give the Contractor notice of that intention*".

Therefore, the employer purported to serve an adequate pay less notice in a timely manner if there is a clear reason to provide less pay; otherwise, a violation of the terms of the contract (i.e., breach of contract) can be made.

3.3.7. Topic 7: Others

Topic 7 has been classified as 'Others' due to its various topics, including legal fees and expense, which are relatively less important than other major topics. The most distinctive terms in Topic 7 are 'cost', 'pay', 'issue', 'fee', 'date', and more specifically, relevant contract provisions have been discovered as follows:

- The parties shall each bear their own legal costs and other expenses incurred in the adjudication;
- Where the referring party is awarded in the aggregate a sum less than 50% of the amount claimed, he shall reimburse the other party the legal costs and other expenses which the non-referring party incurred in the adjudication.

This topic is not related to the cause of the dispute, but rather to the contractual outcome itself resulting from disputes. However, some disputes are associated with legal costs and expenses in the collected legal cases, showing that how to allocate the legal fees and expenses should be clearly defined in advance in the course of construction disputes.

4. Conclusions and Discussion

This study offers domain-specific insights by analyzing construction legal cases and investigating corresponding contractual issues of real project cases to understand possible contract risks that can potentially lead to legal disputes. This study analyzed legal dispute cases by frequency of terms, time periods, and topics. By analyzing the collected documents in terms of term frequency and word correlations, we identified that terms and conditions of the contract are frequently quoted in the relevant legal decisions. Thus, knowing which terms and conditions of construction contracts affect the final legal decision is essential to prevent future disputes. We focused on contract-related content and a judicial decision-related content in the collected cases for contractual knowledge discovery. The main findings of this study are as follows: (1) similar patterns of disputes have occurred repeatedly in construction-related legal cases over the entire period (1999 to 2019), which shows that the factors that have caused construction disputes in the past are being repeated today. Therefore, learning from past failures is meaningful in solving today's problems related to construction disputes. Additionally, the results demonstrate that a large amount of accumulated legal cases is available to solve similar cases, ensuring the feasibility of knowledge discovery using text analytics methods in construction dispute management; and (2) discovered dispute topics indicate that mutually agreed upon contract conditions are important in dispute resolution, thereby demonstrating that in-depth review of contract conditions should be performed in the contracting process. For example, if a certain contract was signed without an expression about the contractor's right to possession of the site, it could be a huge risk for contractors because it could cause construction delay and further conflicts. Such disputes related to possession of the site were observed in the topic modeling analysis. Being aware of contract terms and conditions that are frequently cited in legal judgements will provide practitioners knowledge of the contract terms that should be carefully reviewed at the contract stage. The findings of the study show that our proposed approach can be used to automatically uncover legal knowledge, thereby assisting decision support in contractual strategy formation.

A construction legal case collection is associated with multiple underlying contractual issues. Therefore, gaining a basic understanding of the frequently referenced contract terms and conditions in the legal cases provides useful insights into which contract conditions must considered carefully when reviewing contracts. This understanding enables contract parties to prepare mutually balanced contracts to the roles and responsibilities of the contracting parties. From a methodological perspective, the main contribution of this paper is to present an automated framework to identify trends of disputes and the corresponding contract issues using massive unstructured data while minimizing human effort. In particular, considering that construction legal cases are continuously accumulating information, it is expected that the methodology presented in this study can be expanded in the future analysis. Moreover, the findings of the study are expected to be widely used in issues associated with contract conflicts and disputes by enhancing the accessibility of legal information that requires professional knowledge. The advanced understanding of contractual legal knowledge will be helpful for avoiding uncertainties when drafting contracts.

However, the findings of this study need to be interpreted in the light of its limitations. Due to the nature of case study research, conclusions were derived from the UK legal cases that were used as a source of data for understanding dispute causes and legal decisions. Given the fact that challenges facing the construction industry worldwide are universal, and the disputes issues that are dealt with throughout the current study are common issues witnessed in many countries, the findings of the study provide a useful methodology that can be replicated with data from various countries. Therefore, it is suggested that further case studies extend to other countries, which could provide broader perspectives and understanding.

References

1. Cakmak, E.; Cakmak, P.I. An Analysis of Causes of Disputes in the Construction Industry Using Analytical Network Process. *Procedia Soc. Behav. Sci.* **2014**, *109*, 183–187. [CrossRef]
2. Cheung, S.-O. Critical factors affecting the use of alternative dispute resolution processes in construction. *Int. J. Proj. Manag.* **1999**, *17*, 189–194. [CrossRef]
3. Lee, J.; Yi, J.-S.; Son, J. Development of Automatic-Extraction Model of Poisonous Clauses in International Construction Contracts Using Rule-Based NLP. *J. Comput. Civ. Eng.* **2019**, *33*, 04019003. [CrossRef]
4. Niu, J.; Issa, R.R.A.; Mutis, I. Taxonomy Development toward the Domain Ontology of Construction Contracts: A Case Study on AIA A201-2007. In *Ontology in the AEC Industry*; American Society of Civil Engineers: Reston, VA, USA, 2015; pp. 217–250. [CrossRef]
5. National Bureau of Standards. *National Construction Contracts and Law Survey 2013*; National Bureau of Standards: Newcastle upon Tyne, UK, 2013.
6. Jagannathan, M.; Delhi, V.S.K. Litigation in Construction Contracts: Literature Review. *J. Leg. Aff. Disput. Resolut. Eng. Constr.* **2020**, *12*, 03119001. [CrossRef]
7. Jervis, B.M.; Levin, P. *Construction Law, Principles and Practice*; McGraw-Hill College: Tokyo, Japan, 1988; ISBN 0070374422.
8. Cheung, S.O.; Yiu, T.W.; Chan, H.W. Exploring the Potential for Predicting Project Dispute Resolution Satisfaction Using Logistic Regression. *J. Constr. Eng. Manag.* **2010**, *136*, 508–517. [CrossRef]
9. *Global Construction Disputes Report 2018: Does the Construction Industry Learn from Its Mistakes?* Arcadis: Amsterdam, The Netherlands, 2018.
10. Cheung, S.O.; Asce, F.; Hoi, K.; Pang, Y. Anatomy of Construction Disputes. *J. Constr. Eng. Manag.* **2013**, *139*, 15–23. [CrossRef]
11. Milner, A. Contract interpretation: Potential for relaxing the exclusionary rule. *Int. J. Law Built Environ.* **2011**, *3*, 205–221. [CrossRef]
12. Semple, C.; Hartman, F.T.; Jergeas, G. Construction Claims and Disputes: Causes and Cost/Time Overruns. *J. Constr. Eng. Manag.* **1994**, *120*, 785–795. [CrossRef]
13. Keane, P.; Sertyesilisik, B.; Ross, A.D. Variations and Change Orders on Construction Projects. *J. Leg. Aff. Disput. Resolut. Eng. Constr.* **2010**, *2*, 89–96. [CrossRef]
14. Pena-Mora, F.A.; Sosa, C.E.; McCone, D.S. *Introduction to Construction Dispute Resolution*; Prentice Hall: Hoboken, NJ, USA, 2003.
15. Hughes, W.; Champion, R.; Murdoch, J. *Construction Contracts: Law and Management*; Routledge: Oxfordshire, UK, 2015; ISBN 1317442288.
16. Liu, J.; Li, H.; Skitmore, M.; Zhang, Y. Experience mining based on case-based reasoning for dispute settlement of international construction projects. *Autom. Constr.* **2019**, *97*, 181–191. [CrossRef]
17. Wyner, A.; Mochales-Palau, R.; Moens, M.-F.; Milward, D. Approaches to text mining arguments from legal cases. In *Semantic Processing of Legal Texts*; Springer: Berlin, Germany, 2010; pp. 60–79.
18. Sertyesilisik, B. Investigation on Particular Contractual Issues in Construction. *J. Leg. Aff. Disput. Resolut. Eng. Constr.* **2010**, *2*, 218–227. [CrossRef]
19. Thomas, H.R.; Smith, G.R.; Mellott, R.E. Interpretation of Construction Contracts. *J. Constr. Eng. Manag.* **1994**, *120*, 321–336. [CrossRef]
20. Barman, A.; Charoenngam, C. Decisional Uncertainties in Construction Projects as a Cause of Disputes and Their Formal Legal Interpretation by the Courts: Review of Legal Cases in the United Kingdom. *J. Leg. Aff. Disput. Resolut. Eng. Constr.* **2017**, *9*, 04517011. [CrossRef]
21. 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]
22. Lau, E. Trust and Distrust in Conflicting Situations: Legal Case Review of Antecedents of Distrust. *J. Leg. Aff. Disput. Resolut. Eng. Constr.* **2011**, *3*, 116–122. [CrossRef]
23. Ding, Z.; Li, Z.; Fan, C. Building energy savings: Analysis of research trends based on text mining. *Autom. Constr.* **2018**, *96*, 398–410. [CrossRef]
24. Yin, X.; Liu, H.; Chen, Y.; Al-Hussein, M. Building information modelling for off-site construction: Review and future directions. *Autom. Constr.* **2019**, *101*, 72–91. [CrossRef]
25. Fenn, P.; Lowe, D.; Speck, C. Conflict and dispute in construction. *Constr. Manag. Econ.* **1997**, *15*, 513–518. [CrossRef]

26. Baker, H.R.; Smith, S.D.; Masterton, G.; Hewlett, B. Failures in Construction: Learning from Everyday Forensic Engineering. In *Forensic Engineering 2018: Forging Forensic Frontiers*; American Society of Civil Engineers Reston: Reston, VA, USA, 2018; pp. 648–658.
27. Mitkus, S.; Mitkus, T. Causes of Conflicts in a Construction Industry: A Communicational Approach. *Eng. Constr. Archit. Manag.* **2014**, *13*, 543. [CrossRef]
28. Abdul-Malak, M.-A.U.; Senan, M.H. Operational Mechanisms and Effectiveness of Adjudication as a Key Step in Construction Dispute Resolution. *J. Leg. Aff. Disput. Resolut. Eng. Constr.* **2020**, *12*, 04519051. [CrossRef]
29. Al Qady, M.; Kandil, A. Concept Relation Extraction from Construction Documents Using Natural Language Processing. *J. Constr. Eng. Manag.* **2010**, *136*, 294–302. [CrossRef]
30. Fan, H.; Li, H. Retrieving similar cases for alternative dispute resolution in construction accidents using text mining techniques. *Autom. Constr.* **2013**, *34*, 85–91. [CrossRef]
31. Jallan, Y.; Brogan, E.; Ashuri, B.; Clevenger, C.M. Application of Natural Language Processing and Text Mining to Identify Patterns in Construction-Defect Litigation Cases. *J. Leg. Aff. Disput. Resolut. Eng. Constr.* **2019**, *11*, 1–6. [CrossRef]
32. Yin, R.K. *Case Study Research: Design and Methods*; Sage: Newcastle upon Tyne, UK, 2009; Volume 5, ISBN 1412960991.
33. Adjudication.co.uk—Adjudicator Nominating Body. Available online: http://www.adjudication.co.uk/ (accessed on 11 March 2019).
34. Silge, J.; Robinson, D. *Text Mining with R: A Tidy Approach*, 1st ed.; Tache, N., Ed.; O'Reilly Media, Inc.: Sebastopol, CA, USA, 2017; ISBN 1491981628.
35. Lee, J.; Ham, Y.; Yi, J.-S.; Son, J. Effective Risk Positioning through Automated Identification of Missing Contract Conditions from the Contractor's Perspective Based on FIDIC Contract Cases. *J. Manag. Eng.* **2020**, *36*, 05020003. [CrossRef]
36. Shabbar, H.; Ullah, F.; Ayub, B.; Thaheem, M.J.; Gabriel, H.F. Empirical Evidence of Extension of Time in Construction Projects. *J. Leg. Aff. Disput. Resolut. Eng. Constr.* **2017**, *9*, 04517008. [CrossRef]
37. Royal Institution of Chartered Surveyors. *Final Account Procedures*, 1st ed.; Royal Institution of Chartered Surveyors (RICS): London, UK, 2015. Available online: www.rics.org (accessed on 4 June 2019).
38. Silge, J.; Robinson, D. tidytext: Text Mining and Analysis Using Tidy Data Principles in R. *J. Open Source Softw.* **2016**, *1*, 37. [CrossRef]
39. Wei, T.; Simko, V. An Introduction to Corrplot Package. 2016. Available online: https://cran.r-project.org/web/packages/corrplot/vignettes/corrplot-intro.html (accessed on 1 August 2021).
40. Goetz, J.C.; Gibson, G.E. Construction Litigation, U.S. General Services Administration, 1980–2004. *J. Leg. Aff. Disput. Resolut. Eng. Constr.* **2009**, *1*, 40–46. [CrossRef]
41. Zack, J.G. *Trends in Construction Claims & Disputes*; 2012. Available online: https://www.cmaanet.org/sites/default/files/2018-04/TRENDSINCONSTRUCTIONCLAIMS%26DISPUTES.pdf (accessed on 15 August 2019).
42. Office for National Statistics. *Construction Output in Great Britain: June 2019*; Office for National Statistics: London, UK, 2019.
43. Chaggar, J. *What is Practical Completion and Partial Possession?* Wright Hassall: Warwick, UK, 2015.
44. Thomas, H.R.; Smith, G.R.; Cummings, D.J. Have I Reached Substantial Completion? *J. Constr. Eng. Manag.* **1995**, *121*, 121–129. [CrossRef]
45. Blei, D.M.; Ng, A.Y.; Jordan, M.I. Latent dirichlet allocation. *J. Mach. Learn. Res.* **2003**, *3*, 993–1022.
46. Chen, Y.; Zhang, H.; Liu, R.; Ye, Z.; Lin, J. Experimental explorations on short text topic mining between LDA and NMF based Schemes. *Knowl.-Based Syst.* **2019**, *163*, 1–13. [CrossRef]
47. Lee, J.; Yi, J.-S. Predicting Project's Uncertainty Risk in the Bidding Process by Integrating Unstructured Text Data and Structured Numerical Data Using Text Mining. *Appl. Sci.* **2017**, *7*, 1141. [CrossRef]
48. Steyvers, M.; Griffiths, T. Probabilistic topic models. *Handb. latent Semant. Anal.* **2007**, *427*, 424–440.
49. Wen, Q.; Qiang, M.; Xia, B.; An, N. Discovering regulatory concerns on bridge management: An author-topic model based approach. *Transp. Policy* **2019**, *75*, 161–170. [CrossRef]
50. Zirn, C.; Stuckenschmidt, H. Multidimensional topic analysis in political texts. *Data Knowl. Eng.* **2014**, *90*, 38–53. [CrossRef]
51. Manning, C.D.; Raghavan, P.; Schütze, H. *Introduction to Information Retrieval*; Cambridge University Press: Cambridge, UK, 2008. [CrossRef]
52. Graham, A. Letters of Intent in Construction, Out-Law. 2011. Available online: https://www.out-law.com/en/topics/projects--construction/construction-contracts/letters-of-intent-in-construction/ (accessed on 6 June 2019).
53. Chappell, D. *The JCT Standard Building Contract 2011: An Explanation and Guide for Busy Practitioners and Students*; John Wiley & Sons: Oxford, UK, 2014. [CrossRef]
54. The Law Library. *Housing Grants, Construction and Regeneration Act 1996*; Sweet & Maxwell Ltd.: Mytholmroyd, UK, 1996.
55. Adjudication Society, Chartered Institute of Arbitrators. *Guidance Note: Construction Contracts and Construction Operations*; Chartered Institute of Arbitrators: London, UK, 2013.
56. Mark, W.S. Payment Notices—Do You Prepare and Issue Them Correctly?—Systech International. 2018. Available online: https://www.systech-int.com/issuing-payment-notices-importance-construction-industry/ (accessed on 10 June 2019).