PREVENTING AND CORRECTING QUALITY ISSUES IN CONSTRUCTION WITH INTELLIGENT KNOWLEDGE MANAGEMENT

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

MURATHAN SAYGILI

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF MASTER OF SCIENCE
IN
CIVIL ENGINEERING

JUNE 2022

Approval of the thesis:

PREVENTING AND CORRECTING QUALITY ISSUES IN CONSTRUCTION WITH INTELLIGENT KNOWLEDGE MANAGEMENT

submitted by **MURATHAN SAYGILI** in partial fulfillment of the requirements for the degree of **Master of Science** in **Civil Engineering, Middle East Technical University** by,

Prof. Dr. Halil Kalıpçılar
Dean, Graduate School of **Natural and Applied Sciences** ________________

Prof. Dr. Erdem Canbay
Head of the Department, **Civil Engineering** ________________

Assoc. Prof. Dr. Onur Behzat Tokdemir
Supervisor, **Civil Engineering, METU** ________________

Examining Committee Members:

Prof. Dr. Rifat Sönmez
Civil Engineering, METU ________________

Assoc. Prof. Dr. Onur Behzat Tokdemir
Civil Engineering, METU ________________

Prof. Dr. İsmail Özgür Yaman
Civil Engineering, METU ________________

Assoc. Prof. Dr. Onur Pekcan
Civil Engineering, METU ________________

Asst. Prof. Dr. Ebru Akış
Civil Engineering, Atilim University ________________

Date: 21.06.2022

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name Last name : Murathan Saygılı

Signature :

ABSTRACT

PREVENTING AND CORRECTING QUALITY ISSUES IN CONSTRUCTION WITH INTELLIGENT KNOWLEDGE MANAGEMENT

Saygılı, Murathan
Master of Science, Civil Engineering
Supervisor : Assoc. Prof. Dr. Onur Behzat Tokdemir

June 2022, 91 pages

The construction industry relies on the successful execution of what is planned. Project cost, schedule, and other performance criteria are set and written on a contract, by which all parties are obligated to abide. Similarly, corporate processes are defined by the procedures of companies. However, projects and corporate processes suffer from various obstacles in practice, some of which arise from certain events that might seem insignificant, causing significant deviations from the plan. These obstacles are quality issues that should be handled by preventing them and correcting the existing ones. The thesis addresses this need by providing a quality management (QM) strategy based on knowledge management (KM). Two research questions search for a systemic approach to prevent inadequate quality and correct nonconformities with KM. After showing the relationship between QM and KM, the research builds an information retrieval mechanism based on the cosine similarity metric and a natural language processing (NLP) model called FastText. The development process is divided into two. The first part enables the effective use of what was previously learned, and the second module deals with correction recommendations in the case of nonconformity. Overall, the study enables the recording, storing, and reusing of tacit information to build a continuous learning and continuous improvement cycle. Processing the information in internal audit findings and lessons learned documents with an NLP model, it is argued that nonconformities can be both fixed and prevented. Ultimately, the thesis aims at a state of zero mistakes in construction through continuous learning and continuous improvement.

Keywords: Construction quality management, Knowledge management, Natural language processing, Continuous learning, Continuous improvement

ÖZ

İNŞAATTAKİ KALİTE PROBLEMLERİNİN AKILLI BİLGİ YÖNETİMİ İLE ÖNLENMESİ VE ÇÖZÜLMESİ

Saygılı, Murathan
Yüksek Lisans, İnşaat Mühendisliği
Tez Yöneticisi: Doç. Dr. Onur Behzat Tokdemir

Haziran 2022, 91 sayfa

İnşaat sektörü, yapılması gereken işlerin doğru ve planlandığı şekilde gerçekleştirilmesine dayanır. Proje maliyeti, süresi ve diğer performans kriterleri belirlenmekte, sözleşmeye dökülmekte ve bu süreçlere dahil olan taraflarca uyulması zorunlu kılınmaktadır. Benzer şekilde, kurumsal süreçler şirketler tarafından prosedürler üzerinden belirlenmektedir. Ancak pratikte, projeler ve kurumsal süreçler çeşitli engellerle karşılaşmaktadır. Zaman zaman önemsiz gibi görünen bu gibi engelleyici faktörler, planlarda önemli derecede sapmalara yol açabilmektedir. Bu engeller, önlenmesi ve düzeltilmesi gereken kalite faktörleridir. Bu tezdeki araştırma, önerdiği bilgi yönetimi (BY) tabanlı kalite yönetimi (KY) stratejisi ile bu ihtiyaca yönelik yapılmıştır. Çalışma iki temel araştırma sorusunu hedef almaktadır. Bu soruların ilki kalite problemlerinin BY ile nasıl önlenebileceğini, diğeri ise hali hazırda karşılaşılmış uygunsuzluk durumlarının benzer bir BY stratejisi ile nasıl düzeltilebileceğini sorgulamaktadır. Çalışmada, BY ile KY arasındaki ilişki gösterildikten sonra, kosinüs benzerlik ölçütü ve FastText adı verilen doğal dil işleme (DDİ) modeli kullanılarak bilgiye erişim mekanizması geliştirilmiştir. Geliştirme süreci iki başlık altında incelenmiştir. Birinci kısım öğrenilmiş derslerin etkili şekilde kullanılmasını sağlarken ikinci kısım ise uygunsuz durumlara uygun aksiyonların alınmasına odaklanmaktadır. Genel hatlarıyla incelendiğinde bu çalışma, örtük bilginin kaydedilmesi, saklanması ve yeniden kullanılmasını sağlayarak sürekli öğrenme ve sürekli iyileştirme döngülerinin geliştirilmesini hedeflemektedir. Araştırma, denetim bulgularının ve öğrenilmiş derslerin DDİ ile işlenerek düzeltilebileceği ve önlenebileceğini önermektedir. Sonuç olarak bu tez, inşaat süreçlerinin sürekli öğrenme ve iyileştirme döngüsü içerisinde sıfır hataya ulaşılmasını amaçlamaktadır.

Anahtar Kelimeler: İnşaat kalite yönetimi, Bilgi yönetimi, Doğal dil işleme, Sürekli öğrenme, Sürekli iyileştirme

Dedicated to my family

ACKNOWLEDGMENTS

I would like to express my gratitude to Assoc. Prof. Dr. Onur Behzat Tokdemir for his priceless guidance and contributions to this study. He has always been a great mentor and advisor in my academic and professional career and personal life.

I wish to show my deepest gratitude to my mother Yasemin Saygılı, my father Murat Saygılı, and my brother Atahan Saygılı. They never stopped believing in me and supported me throughout my entire life, and there are not enough words to express how lucky and proud I am to be a part of this family.

I would also like to thank Bilal Umut Ayhan for his help during this study. He always enlightened me with his ideas and recommendations, and I learned a great deal of knowledge from his research.

I want to thank the examining committee members for their valuable feedback and contributions to this study. Their ideas enlightened not only this thesis but also the future work.

Finally, I would like to thank Su Yapı Engineering and Consulting, Inc. and ENKA Construction and Industry, Inc. for their contributions to my graduate study. I gained a wealth of experience with the help of people working at these companies.

TABLE OF CONTENTS

LIST OF TABLES

TABLES

LIST OF FIGURES

FIGURES

LIST OF ABBREVIATIONS

ABBREVIATIONS

AI	Artificial Intelligence
API	Application Programming Interface
CBR	Case Based Reasoning
ISO	International Organization for Standardization
IT	Information Technologies
ITP	Inspection and Test Plans
KB	Knowledge Base
KM	Knowledge Management
KMS	Knowledge Management Systems
ML	Machine Learning
NCR	Nonconformity Report
NLP	Natural Language Processing
QA	Quality Assurance
QC	Quality Control
QM	Quality Management
QMS	Quality Management Systems

CHAPTER 1

INTRODUCTION

The world is evolving without rest, and the construction industry is no different. Customer demands, construction techniques, standards, resources, and many more are changing while transforming the industry. The traditional view of the industry regarding project budget and schedule was added new concepts such as quality and sustainability. Although the well-known iron triangle of the industry shows quality as a fundamental part of construction projects (Larsen et al., 2016), the reality tends to omit it due to time and cost limitations. The idea is either a shorter time or a lower budget while planning and designing construction projects. A common strategy is that companies should spend more money and allocate more resources to projects if they aim to reduce the total project duration. Similarly, if they need to limit the project budget, construction companies should lower their capacity, which leads to an increase in the project duration. Undoubtedly, the time-cost tradeoff applies successfully to construction projects. However, there is a third leg, quality, when the true success of construction projects is considered.

Prioritizing quality can not only prevent an unexpected increase in the project duration but also keep the project budget within desirable limits. Quality in this context is both the value of the delivered product and business processes and their compliance with standards and other regulatory bodies (Arditi & Murat Gunaydin, 1997). With improved quality, the rate of rework and waste can be lowered, resulting in reduced cost and duration. This waste can be the waste of material, excessive time and cost, idle effort, or loss of knowledge. When projects deviate from their quality objectives, an increase in project cost and duration harms project stakeholders. Burati Jr. et al. (1992) stated that quality deviations raise the project budget by approximately 12.4%. Therefore, making decisions and taking actions to improve quality should be at the core of the business strategies of construction companies. Setting and implementing a quality management system (QMS) is a way to reduce waste and improve quality.

A well-designed and implemented QMS can elevate the quality level to its desired state (Battikha, 2002). QMS primarily consists of quality assurance (QA) and quality control (QC). QA ensures that quality requirements are fulfilled in projects, whereas QC is the actions that lead to QA goals (Arditi & Murat Gunaydin, 1997; Burati Jr. et al., 1992). Guidelines and standards such as ISO 9000 series (International Organization for Standardization, 2015) are used by construction companies to achieve the goal of sufficient quality. According to ISO 9000 standards, quality management (QM) principles include process approach, improvement, and evidence-based decision-making. Firstly, for a system to succeed in the area it aims, it should be integrated into the existing systems and processes to increase effectiveness and efficiency, which turns into an overall rise in organizations' performance. Secondly, there should be a continuous improvement in the system. The construction industry and its projects are dynamic, going through frequent changes. Therefore, the system should adapt to these changes. Finally, the system should be evidence-based, which necessitates supporting the entire process with previous documentation and experience. When these practices are not followed, lack of quality tends to emerge as a core problem for construction companies.

It is widely known that the construction industry is infamous for poor performance and quality (Ali & Rahmat, 2010). This insufficiency has impacts on both project-level and corporate-level. Nonconformance may frequently occur in construction projects when the deliverables, either an entire project or its subunits, are not compliant with the predefined set of rules and quality levels (Battikha, 2002). The cases when a required quality level is not achieved are recorded in nonconformity reports (NCR). Similarly, office functions are also affected by nonconforming processes. These processes are periodically audited, comparing existing actions with the standards. In both practices, findings are recorded, as well as root causes, corrections applied, and preventive measures taken. Furthermore, to minimize or prevent the negative impact of poor quality, construction companies take certain measures. In the early phases of construction projects, inspection and test plans (ITP) are prepared as part of QA strategies. These documents include requirements and acceptance criteria, frequency and scope of inspections and tests, and parties responsible for their approval (Battikha, 2002). However, in spite of these practices, lack of quality in the construction items and processes may lead to an additional financial burden on the project budget, unexpected delays, safety issues, and many more (Zulkifli et al., 2018).

Various factors are challenging the implementation of QMS and causing the aforementioned adverse effects. One challenge is following a paper-based approach during site inspection (Zulkifli et al., 2018). In such practices, missing and inconsistent information lower the effectiveness of QM strategies, and they require significant effort and time. Other challenges are mentioned by (Arditi & Murat Gunaydin, 1997). Lack of top-management commitment and absence of feedback mechanism are among the factors that hinder effective QM practices. Addressing these challenges can fix the industry's attitude towards quality. Developed solutions should minimize nonconformance, provide effective corrections to the ones that occurred, identify the root causes, and ensure that such nonconformance will not recur (Battikha, 2002). One of the solutions having the ability to address these is proper knowledge management (KM) since effective use of knowledge can prevent quality problems and increase stakeholder satisfaction (Love et al., 2003). Therefore, managing the vast amount of knowledge is crucial for construction QM. An intelligent system that stores and carries this knowledge further is necessary so that the probability of encountering misconducted work processes can be reduced while increasing value, quality, and stakeholder satisfaction in return.

As a result, the main objective of this thesis is to adopt KM by leveraging state-of-the-art natural language processing models (NLP) strategies to lower mistakes and improve quality while creating a continuous learning and continuous improvement cycle (Figure 1.1.). With this approach, it is aimed that knowledge can help improve the QM processes in construction, leading to a state of zero-mistake. The study was raised due to QM challenges and lacking KM practices, elaborated further in the following chapters. The primary point of departure of this study is that current KM strategies are incompetent to process collected information that could be used in future cases to prevent or correct nonconformance. When the recorded information is not wasted and transferred into future processes, KM fosters successful QM applications.

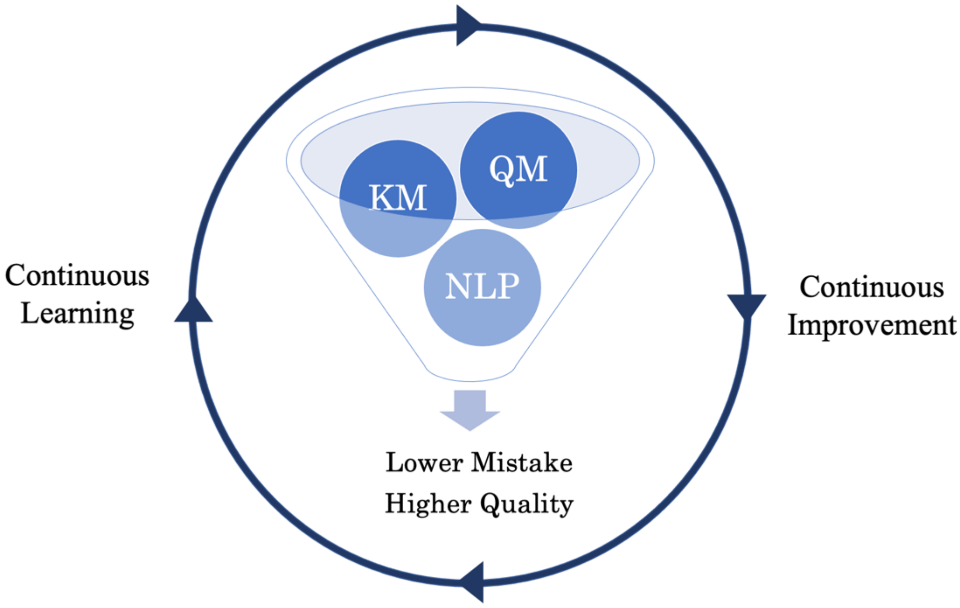

Figure 1.1. Research Objective

The thesis processes knowledge obtained and stored in various documents that are the output of internal QM processes of a large construction company (referred to as the Company). The Company ranked among the top 100 of the "Top 250 International Contractors" list of Engineering News-Record (ENR). Their QMS guided this study and provided insightful information. As a result, the following research questions shaped the form of this study.

(i) How can the construction industry prevent quality problems using knowledge management in a continuous improvement cycle?

(ii) How can the construction industry improve the correction process of nonconformities by using knowledge management in a continuous learning cycle?

The proposed framework and the developed tool aim to improve and assist the existing QMS. The thesis is structured as follows. The following chapter provides a background on KM and NLP. Then, Chapter 3 reviews the literature on QM, KM, and NLP studies in the construction research area. Chapter 4 emphasizes the relationship between KM and QM by providing real data from the industry and taking support from the literature. In Chapter 5, the research methodology is explained, and how the NLP model was developed is detailed in Chapter 6. Integration and implementation of the proposed approach into current practices is explained in Chapter 7 by merging the tool with the existing QM strategies of the Company. Also, three implementation scenarios are given in this chapter. Discussion and evaluation are provided in Chapter-8, and final comments and conclusions are provided in Chapter 9.

CHAPTER 2

BACKGROUND ON KNOWLEDGE MANAGEMENT AND NATURAL LANGUAGE PROCESSING

2.1 Knowledge Management

Knowledge has been one of the most valuable resources for construction companies as construction projects are driven by the experience of project managers, engineers, supervisors, workers, and so forth (P. Carrillo, 2004a; Pheng & Chuan, 2006; Tserng & Lin, 2004). This knowledge can be decisions, actions, failures, solutions, preventive measures, and best practices. With sufficient experience and know-how, companies could raise the value of their tangible assets.

After emerging of large corporations, the definition of assets has shifted into something more intangible; the information (Z. Ding et al., 2007) and the knowledge itself became an asset for those who carry the industry forward. Leading firms decided to take action to keep up with their competitors in terms of this new kind of asset, and they started to collect data and store it in their databases. However, what is as crucial as collecting and storing that data is knowing what to do with it. Managers and employees should be aware of the potential of the resource in their hands. Thus, processing the stored data to extract information is necessary to generate knowledge from it so that better decisions can be taken (Choo, 2006). These discussions and practices require a thorough understanding of KM and its functions.

KM incorporates the processes that create, secure, collect, coordinate, combine, retrieve and share knowledge (Tserng & Lin, 2004). The knowledge can be tacit or explicit; however, effective knowledge management systems (KMS) should take both types into account. Tacit knowledge is defined as the type of knowledge that is difficult to codify in a formal language (Pathirage et al., 2007), whereas explicit knowledge can be represented and processed easily (Burton et al., 2011; Tserng & Lin, 2004). The main function of KM is to turn that knowledge into an asset, regardless of its source or tacit nature (Robinson et al., 2004). For instance, such an asset can be the lessons learned in the construction industry (T. Kim & Chi, 2019). Recording lessons learned is essential, but processing that knowledge is what actually creates a continuous learning and continuous improvement culture (Robinson et al., 2005). Therefore, the knowledge should be maintained appropriately and reused effectively as they are important not only for ongoing projects but also for upcoming ones.

Although it has primary importance, managing knowledge has not been an easy task for companies. Improvement means change, which organizational culture tends to resist (C. Egbu et al., 2001; Robinson et al., 2004). Therefore, effective implementation of KM practices should expect to face barriers (Robinson et al., 2005). First of all, construction firms do not intend to lose their competitive advantage over others; therefore, they act reluctant to information sharing. Even in the internal dissemination of information within a single company, employees are not willing to give away their experience to their colleagues because of hierarchical structure and competitiveness (Z. Ding et al., 2007). In addition, time limitations, lacking integration of KM into existing strategies, and capturing difficulties of information also reinforce the challenges that are faced while setting and implementing KM practices (*Knowledge Management Research Report*, 2000).

Furthermore, Tupenaite et al. (2008) stated that both tacit and explicit knowledge leads to problems and challenges. In the case of tacit information, the fragmented nature of the industry and the temporary nature of construction projects reinforce the challenges that make transferring tacit experience and knowledge difficult (Carrillo, 2004). Personnel changes from process to process or from project to project are one of the main reasons for this challenge since current practices carry the knowledge mainly through employees that were involved in previous projects. Project managers specialize in certain areas and prevail among other employees due to their experience. However, this approach is prone to human error. In the worst case, when an experienced employee leaves the firm, knowledge, experience, know-how, and problem-solving strategies become entirely lost. Therefore, the industry is highly fragile to employee turnover due to tacit knowledge.

In addition, explicit knowledge faces issues in information recording in that partial or incomplete records are abundant in the industry. This is also one of the challenges of this thesis study. Although the study was fed with a relatively large knowledge base, missing information challenged the progress of the thesis. However, Tupenaite et al. (2008) also recommended the effective use of IT applications and codification of tacit knowledge to overcome these challenges, which this study aimed to achieve. Finally, to build a well-functioning KMS for construction companies and implement it successfully, top-level management should understand the need and encourage the development of this system; otherwise, lack of administrative support also hampers KM practices (Robinson et al., 2005). Table 2.1. expands and summarizes the challenges faced while implementing KM practices in the construction industry.

Table 2.1. Challenges in front of KMS in Construction

Challenges	References
Organizational culture	(P. Carrillo et al., 2004; Z. Ding et al., 2007; Hallowell, 2012; Robinson et al., 2004; Sik-wah Fong & Chu, 2006)
Poor communication	(Robinson et al., 2004; Sik-wah Fong & Chu, 2006)
Resistance to sharing of knowledge	(P. Carrillo et al., 2004; P. M. Carrillo et al., 2002; Z. Ding et al., 2007; Hallowell, 2012; Robinson et al., 2004, 2005; Sik-wah Fong & Chu, 2006)
Initiative overload	(Robinson et al., 2004, 2005)
Poor IT infrastructure	(P. Carrillo et al., 2004; Hallowell, 2012; Robinson et al., 2004, 2005)
Lacking executive support	(Robinson et al., 2004, 2005; Sik-wah Fong & Chu, 2006; Tupenaite et al., 2008)
Conflicts of KM and other business functions	(Robinson et al., 2004, 2005)
Absence of standardized work processes	(P. Carrillo et al., 2004)
Fragmented nature	(P. Carrillo et al., 2004)
Time restrictions	(P. Carrillo et al., 2004; P. M. Carrillo et al., 2002; Hallowell, 2012; *Knowledge Management Research Report*, 2000)
Budget restrictions	(P. Carrillo et al., 2004; Hallowell, 2012; Robinson et al., 2005)
High tacit nature	(Tupenaite et al., 2008)
High turnover rate	(Hallowell, 2012; Tupenaite et al., 2008)
Missing records	(Tupenaite et al., 2008)

Despite these challenges, there are attempts to manage knowledge in the industry. However, most of the current approach to knowledge is quite inadequate. Most construction firms are not aware of the importance and potential of data, which is the source of information and knowledge. Even if they acknowledge the benefits of KM, expecting a return in the short-term results in failure of KMS since KM adds value in the long run (Hallowell, 2012). On the contrary, academia has had a strong interest in KM for decades. Nevertheless, the ones taking advantage of state-of-the-art techniques and technologies are rare. Most research in the literature is not data-driven and automated because of the high tacit nature of information required by construction (T. Kim & Chi, 2019; Zou et al., 2017). These result in time losses and further inefficiencies in KM processes (T. Kim & Chi, 2019). As people in the industry cannot break the fear factor raised by the idea that information processing requires excessive effort, companies stay reluctant to disrupt their workflows. However, current technologies and techniques enable the effortless processing of data, which can help the development of effective and efficient KM strategies in the construction industry.

Since most of the data generated and used by construction companies are stored in a text format (Baek et al., 2021; Caldas et al., 2002), text analysis and text mining tools are crucial to achieving a successful outcome in research on KM. Since tacit knowledge is hard to store in a structured format, unstructured text format is a common way to store it. Also, the recent literature on KM mostly focused on knowledge use and exploitation, its transfer, and utilization of information and communication technologies (Yepes & López, 2021). Therefore, NLP, a widely used technique for text mining, can be adopted in this research area so that knowledge exploitation and its dissemination is improved.

2.2 Natural Language Processing

NLP is one of the application fields of artificial intelligence (AI) that helps process unstructured text data so that computers can understand it (Candaş & Tokdemir, 2022a; Jallan & Ashuri, 2020; Lee et al., 2019). It can be used in various cases, including information extraction, information retrieval, named-entity recognition, machine translation, sentimental analysis, document classification, and question answering (Grzegorczyk, 2018; Shin & Issa, 2021). The underlying methodology in research using NLP can be categorized into three main groups which are lexical, syntactic, and semantic analyses. Lexical analyses perform a word-level study. On the other hand, in a syntactic search, word sequences are taken into account, so it is closer to a sentence-level analysis. For instance, understanding phrases and grammar falls into this group. Finally, the ultimate search, the meaning of words and sentences, is performed by a semantic analysis. Here, both morphological and contextual outcomes are provided to researchers (Lee et al., 2019).

In NLP studies, feature extraction from text is important as these features represent the words and sentences (Kowsher et al., 2022). Converting text into vectors is used for this purpose, and this process is called word embedding (Dharma et al., 2022). The selection of a proper word embedding model is crucial in NLP research as the performance of studies depends on selecting a proper model. There are several word embedding algorithms widely used in the natural language field. Word2Vec, GloVe, and FastText are such examples. These methods are preferred in most research as they can capture semantic and syntactic structure in sentences (Dharma et al., 2022). In this study, FastText was used as the word embedding model.

FastText is an open-source library developed by Facebook AI Research (FAIR) Lab in 2016 (Bojanowski et al., 2016; *FastText*, 2020; *FastText*, 2022; Grave, 2016; Mikolov et al., 2016). It was mainly built for text classification and text representation to address the need for a better understanding of large text data. The algorithm was developed after making improvements to the vanilla word2vec model (Dharma et al., 2022; Jallan & Ashuri, 2020). FastText represents text data with the bag of n-grams and subword information rather than only using the skip-gram model. It also uses hierarchical softmax to the lower execution time of computation. Unlike word embedding algorithms utilizing the bag of words model, FastText takes word order into account. For instance, it was stated by developers (Mikolov, Chen, et al.,

2013) that FastText has a better performance than word2vec (Mikolov, Chen, et al., 2013). The model's performance assessment results are provided in Table 2.2. The table shows that it competes with deep learning-based algorithms with a slight compromise on the accuracy, and it dramatically outperforms those algorithms in terms of time. In addition, (Dharma et al., 2022) compared Word2Vec, GloVe, and FastText and revealed that FastText prevailed among others in terms of accuracy. These findings are also provided in Table 2.3. Therefore, it can be seen that FastText reduces computation time significantly, enabling faster research. Furthermore, the FastText library consists of word vectors for 157 languages, including Turkish. Having such variety enriches the reasons why the library was preferred in this study.

Table 2.2. Performance of FastText and Deep Learning-Based Models (Mikolov et al., 2016)

	Yahoo		**Amazon full**		**Amazon polarity**	
	Accuracy	**Time**	**Accuracy**	**Time**	**Accuracy**	**Time**
Char-CNN	71.2%	1 day	59.5%	5 days	94.5%	5 days
VDCNN	73.4%	2 hours	63.0%	7 hours	95.7%	7 hours
FastText	72.3%	5 sec.	60.2%	9 sec.	94.6%	10 sec.

Table 2.3. Accuracy Scores of Different Word Embedding Models (Dharma et al., 2022)

Word Embedding Model	**Accuracy (%)**
Word2Vec	92.5
Glove	95.8
FastText	97.2

CHAPTER 3

LITERATURE REVIEW

The thesis was built on three main topics and their use in the construction industry. These topics are QM, KM, and NLP. Therefore, the literature was reviewed under these three categories.

3.1 Construction Quality Management

There is extensive literature on construction QM. Therefore, this section focuses on the construction QM from an AI perspective. However, firstly, it is important to understand the increase in overall interest in construction QM. Therefore, the research trends on this subject are given in Figure 3.1.

The data feeding Figure 3.1. was collected from the Scopus database on March 4th, 2022. The search keywords were 'Construction Industry' and 'Quality Management'. The results were filtered; only the construction-related research fields remain. Also, to observe a complete trend, research published in 2022 was excluded from the graph. As a result, research regarding QM in the construction industry since 1969 can be observed. The bars on the graph showed that QM in construction had been steadily gaining interest. Moreover, the graph also shows the QM studies adopting AI, shown with the blue line. Despite a fluctuating history, the AI-related quality management research started to gain attention after 2000 and showed a dramatic increase after 2017. This result showed that it is a field with high demand for research and innovation. Hence, a summary of the innovative approach to the topic is provided in this section of the thesis.

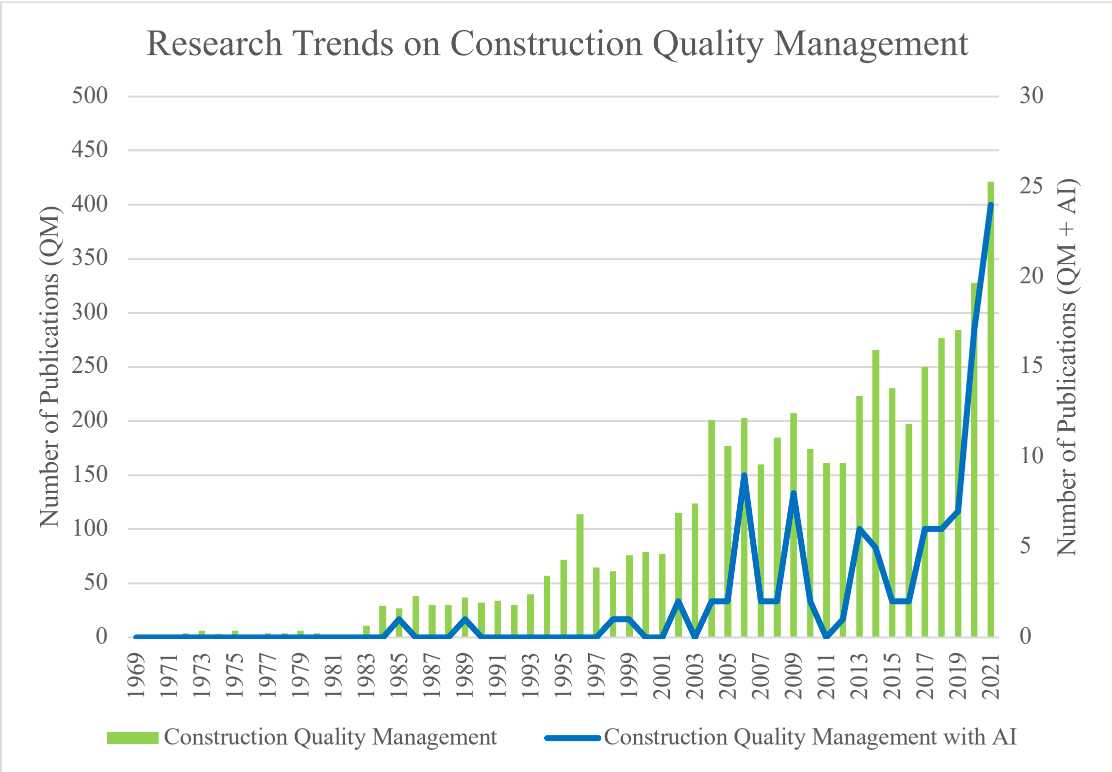

Figure 3.1. Research Trends on Construction QM Between 1969 and 2021

QM studies adopting AI are limited in the literature, and the majority of related studies were published after 2017. A study that belongs to the early 2000s Soibelman and Kim (2002) delved into the preparation of data to generate knowledge in the construction industry. They revealed the crucial steps while adopting contemporary technologies like AI and machine learning (ML), and they used their model to determine the causes of delays, cost overrun, and issues in QA/QC. After 2010, Asgari and Rahimian (2017) reviewed the tools and technologies that adopted AI and proposed a conceptual monitoring framework for the identification, reporting, and documentation of construction defects. These studies and others that are in the scope of this research paved the way toward more advanced research in this area.

At the time when research interest in AI for QM skyrocketed, Fan et al. (2021) conducted research regarding ML applications on reinforced concrete bridges. They studied various machine learning methods in the QM of reinforced concrete bridges and their inspections. Besides the structural perspective, Egwim et al. (2021)

investigated the factors leading to project delays in the construction industry of Nigeria. The results showed that QC was found to be the most important factor in the delay of construction projects. They also suggested proper use of data obtained from schedule backlogs, as-built drawings, costs, invoices, employee information, etc. According to their statements, processing the data to benefit from it is possible with the application of AI and ML. Liu et al. (2022) had their focus on modular construction and the application of AI-based object detection in this field, and they emphasized the importance of object detection for quality and safety in construction projects. They compared two deep learning models (faster region-based convolutional neural network – faster RCNN and single-shot multi-box detector – SSD). In terms of the metrics they used, average recall and average precision assessed the model outcomes. Recently, Zhu et al. (2022) studied the use of smart technologies in schedule, budget, and quality management areas in construction project management and provided a review on the smart applications in project QM. They stated that AI applications could improve project performance in the construction industry.

3.2 Knowledge Management In Construction

KM practices were also widely adopted by construction companies. These practices were used in both corporate and project-based processes. Despite the steady increase in interest in the construction QM, KM-related research in the construction domain has been gaining interest exponentially, as shown by the green bars in Figure 3.2. When AI was included in the search, the blue line reveals that there has been strongly fluctuating behavior in the research focus. However, what is important is the recent rise in the number of research. Similar to Figure 3.1., Figure-3.2. supports the argument that the application of AI in the construction KM domain is still needed. This section summarizes the research on construction KM.

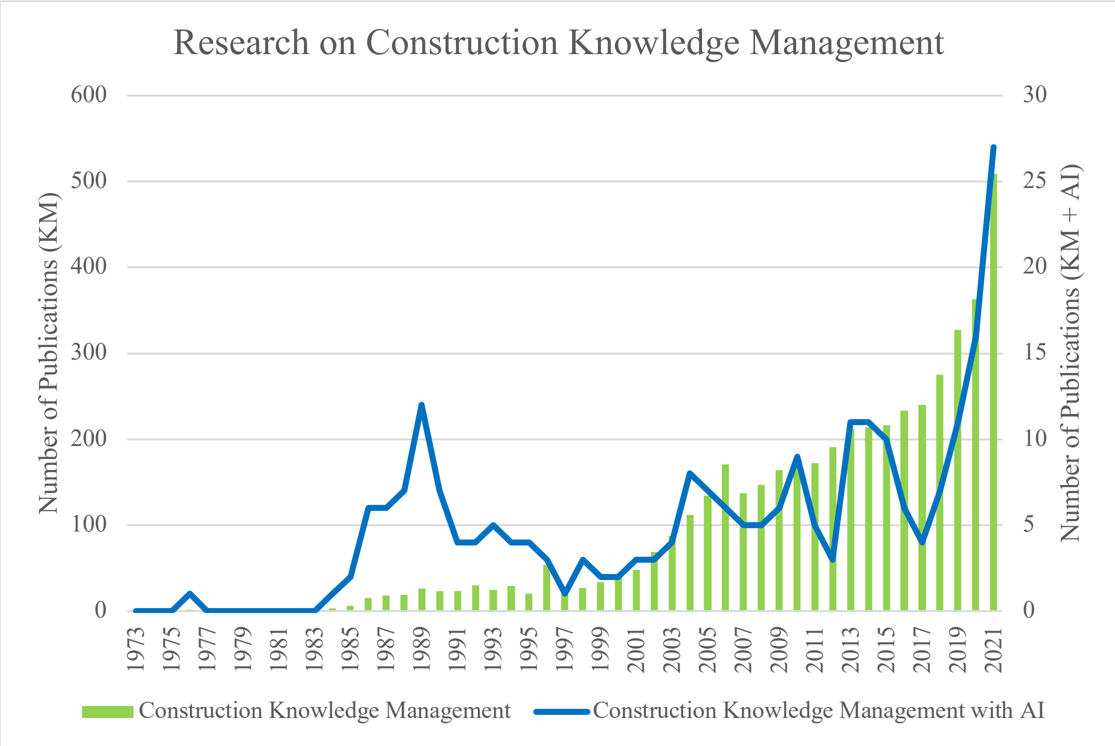

Figure 3.2. Research Trends on Construction KM Between 1969 and 2021

Both universities and companies delved into KM practices to leverage their preexisting knowledge. The initial group of studies revealed the need for and importance of information and knowledge for the industry. Carrillo (2004) revealed the primary reasons behind such research by looking into the oil and gas industry. According to the findings, the three of the most observed reasons were the necessity of continuous improvement, communicating best practices, and reducing rework. Moreover, Pathirage et al. (2007) focused on the importance of tacit information for the construction industry and revealed the benefits of using tacit knowledge in terms of organizational performance. Robinson et al. (2004) revealed why companies need effective KM. These needs were communicating the tacit information in the minds of key personnel, spreading the knowledge and the use of best practices, and decreasing rework. However, the ultimate motivation was to achieve better business performance and higher profit.

In a business-focused study, Robinson et al. (2004) used a framework called IMPaKT to reveal the strategic importance of knowledge and its management, develop a KMS to achieve business objectives, and create a benchmark for knowledge management in terms of business performance. Styhre and Gluch (2010) observed the use of platforms that enable standardized methodology for managing and sharing knowledge in a major Scandinavian construction company. Finally, Adenfelt (2010) delved into the effect of knowledge sharing on the performance of transnational projects. The research showed that low communication and coordination have a negative impact on transnational project performance. As it was stated by previous research that KM can increase communication and collaboration, the importance of this field of study becomes obvious.

The importance of knowledge made it necessary to select a proper KMS. Kamara et al. (2002) studied the selection of the most appropriate KMS for organizations. Their research focused on construction and manufacturing industries and observed various knowledge management practices in them. This selection might be crucial as there are various identified factors that have an impact on effective KM and obstacles that hinder it. Sik-wah Fong and Chu (2006) studied the obstacles, critical factors, and gains regarding the practice of effective knowledge sharing in tendering departments of construction companies in Hong Kong and the United Kingdom. In the following year, Ding et al. (2007) focused on the reasons behind effective knowledge sharing and trust within architecture firms in China, stating that employees' characteristics are the primary reason for that. According to the results of their study, employees' attitude, ability, personality, and social interaction with others were the four main reasons for this issue. They also stated that interpersonal trust does not have a significant effect on knowledge sharing.

There were also studies searching for the existing KMS. P. Carrillo et al. (2004) revealed KM strategies, resources, and challenges in the UK by conducting a survey. Following the research, Robinson et al. (2005) studied large construction companies in the UK to observe how they apply KM strategies and developed a framework called STEPS for KM maturity assessment for companies. Also, Soibelman and Kim

(2002) revealed the five necessary steps for construction knowledge discovery. These steps were identification of problems, data preparation, data mining, data analysis, and refinement. In addition to these strategies, there were tools developed for KM purposes. Tserng and Lin (2004) focused on KM applications in the construction phase of projects and developed a system called Construction Activity-Based Knowledge Management (ConABKM). They validated the effectiveness of KM in the construction industry by implementing their prototype in a highway project. Recently, Eken et al. (2020) focused on organizational learning and developed a lesson learned management process model to collect and share construction knowledge.

When the importance of proper management of knowledge was understood, KMS was applied to various areas of the construction industry. One of these areas was related to document management and information retrieval. Meziane and Rezgui (2004) developed a method that can be used for the organization, storage, and retrieval of construction-related documents based on their similarity. Also, H. Kim et al. (2015) sought an automated information retrieval system and proposed a retrieval framework that used BIM objects as query sets so that the new set, i.e., the new case, can be compared to previous ones according to a similarity index. Based on survey results they conducted, they concluded that such a system could be effective in preventing construction accidents.

Another use case was safety management. Two consecutive studies (Goh & Chua, 2009, 2010) proposed a method to identify construction hazards. The study benefited from the case based reasoning (CBR) approach by codifying and retrieving past hazard cases depending on the current need. They also followed the job hazard analysis approach in their study. Finally, they used similarity-based scoring while retrieving past cases. Cambraia et al. (2010) suggested instructions to identify, analyze, and share information regarding near-miss accidents on construction projects. When they applied what they suggested in a healthcare building project, they surfaced accident cases that were previously overlooked. Hallowell (2012) also focused on construction safety and searched for the safety-related knowledge management methodologies adopted by the construction industry through 11 case studies. They concluded that the industry lacks in storing and transferring knowledge and therefore hinders effective safety-related corrections. Lu et al. (2013) used the CBR approach to identify safety risks on subway projects and recommend preventive measures to avoid potential accidents. Lastly, Ayhan and Tokdemir (2019) adopted CBR with artificial neural network technique to predict safety-related scenarios and preventive safety incidents.

Other studies analyzed the KM-IT duo. Tupenaite et al. (2008) provided different views on IT and KM applications in the literature and revealed the benefits of their adoption in the construction industry. Their study developed a model following a four-step framework; gathering knowledge, collecting knowledge, creating best practice data, and providing decision support in upcoming processes. Yang et al. (2012) aimed at revealing the relationship between KM, project performance, and IT practices. They stated that KM could help achieve project goals. Similarly, P. M. Carrillo et al. (2002) examined the use of IT within the context of construction KM.

3.3 NLP Applications In The Construction Industry

Although NLP use is relatively new to the construction industry, there are several use cases published in the literature. These application areas are contract management, safety, building information modeling, staff assignment, document management, risk management, and scheduling.

As for KM applications in contract management, Al Qady and Kandil (2009) aimed at accessible and efficient KM for project and contract management activities and used NLP for information retrieval in order to achieve this goal. They developed a technique, concept relation identification using shallow parsing (CRISP), and used it to enhance electronic document management applications with the help of extracted semantic knowledge from construction documents (Al Qady & Kandil, 2010). With a similar interest, Zhang and El-Gohary (2012a) and Zhang and El-

Gohary (2012b) observed the effectiveness of syntactic and semantic NLP approaches for the automated extraction of regulatory information existing in building codes. Jallan et al. (2019) automated the construction-defect litigation analysis process using NLP. Lee et al. (2019) developed an NLP-based automated model that extracts risk information from contract clauses in construction projects. Their aim was to support the review process of contracts so that claims and disputes could be prevented.

Recently, Candaş and Tokdemir (2022a) proposed a methodology to identify vague sentences in construction contracts using NLP. The corpus they used to train their model was construction contracts prepared by the International Federation of Consulting Engineers (FIDIC). They observed reduced time in the contract review process, high accuracy, and less dependence on expert opinion. The model they developed used Natural Language Toolkit (NLTK) library created by Bird et al. (2009), and they combined the bag-of-words model with the term frequency-inverse document frequency (TF-IDF) function in their research. In the following study, the authors (Candaş & Tokdemir, 2022b) developed a multilabel text classification tool to identify relevant departments for the contract review process. They aimed at making contract reviews more efficient and effortless. From a deep learning perspective, Moon et al. (2022) conducted research on construction specifications and proposed an automated document review using NLP. Their study was composed of three layers. The first one focused on semantic properties of specification documents and created a thesaurus for construction-related terms, which was enabled by the word2vec model. In the second step, they used named-entity recognition with bi-directional LSTM. The final step developed a provision-pairing model by means of the Doc2Vec embedding model and identified the most relevant provisions.

When safety was the concern, Tixier et al. (2016) delved into construction safety and adopted NLP to analyze the contents of accident reports. In a similar study, T. Kim and Chi (2019) retrieved and analyzed accident cases in construction projects using NLP techniques. They developed a KM tool with an NLP-based technique that processed the unstructured text data of safety accidents. Their research was divided into two in that the first part was for information retrieval, and the second part focused on analysis. In the first part, they utilized query expansion using a construction thesaurus and Okapi BM25 for ranking the data. The second part used rule-based and conditional random field (CRF) techniques to achieve the research goal. Chokor et al. (2016) based their research on unsupervised ML and NLP to analyze injury reports of the Occupational Safety and Health Administration (OSHA) in Arizona. During the preprocessing phase, they used three techniques to prepare their unstructured data. These methods were removing stopwords, stemming, and tokenizing. Furthermore, they used TF-IDF for weighting terms and K-means clustering for unsupervised ML.

Building information modeling (BIM) also made use of KM strategies. Salama and El-Gohary (2011) proposed an automated regulatory compliance checking strategy for construction projects and building information models. Their primary aim was to overcome the challenge of the complexity of laws and regulations while interpreting the meaning of sentences, which had been modeled using if-then rules. Their research comprised a five-step study, including the extraction and formalization of rules in contracts and other project documents, representing the text using NLP, and preparation of BIM models. They also referred to deontology, the theory of obligations and rights, and deontic logic. Shin and Issa (2021) developed a framework called Building Information Modelling Automatic Speech Recognition (BIMASR) so that users can search and manipulate the model through speaking without any prior knowledge of BIM commands. They provided two case studies with a single-object BIM model and a multi-object BIM model to verify the framework they proposed. Recently, Wang et al. (2022) developed a query-answering tool for information extraction from BIM. They achieved building a virtual assistant by using NLP. Their system was composed of natural language understanding, information extraction, and natural language generation. They utilized the NLTK library of Python language. Also, the researchers adopted term frequency (TF) using the bag-of-words (BOW) method for vectorization and cosine similarity for information extraction.

While assigning the right personnel to the right tasks, studies used NLP. Bafna et al. (2019) proposed an automated task recommender system for both personnel and task allocation processes. They used TF-IDF to convert the unstructured text data into a structured vector format. Their methodology depended on the candidates' skillsets that were extracted from their resumes. Therefore, they adopted knowledge management practices in their research, as well. Similarly, Mo et al. (2020) addressed the challenge of staff assignment in building maintenance and proposed an NLP model that automatically assigns staff and priority based on an existing request.

Document classification was among the research interest of scholars of the construction industry. Caldas et al. (2002) aimed to improve information management in construction with an automated document classifier and provided a prototype of such tool. Other studies focused on the status-quo of the construction industry by taking advantage of NLP tools. Tang et al. (2020) aimed at discovering the similar and different conditions in the construction sectors of China and the United States (US). They used data collected from the social media platforms of two countries and processed that data via big data analytics tools. They utilized StanfordNLP (Manning et al., 2014) for sentiment analysis to categorize shared posts as positive, negative, or neutral. The results showed that the users from the US mentioned safety and energy topics more than the ones residing in China. Recently, Baek et al. (2021) provided an in-depth review of text-based research in the construction literature. They studied various data sources and text analysis methods. Similarly, Y. Ding et al. (2022) reviewed data sources, tools, and technologies in the field of NLP in the construction sector.

There were studies focusing on risk management, scheduling, and monitoring in construction. Zou et al. (2017) followed a case-based reasoning approach to retrieve risk cases in the construction industry, and they adopted NLP to analyze text. Erfani et al. (2021) evaluated the project risk similarity using NLP. They used Word2Vec on risk register documents. Furthermore, Amer et al. (2022) focused on scheduling. Their Long-Short-Term-Memory (LSTM)-based model analyzed schedule patterns and learned activity relations to improve construction schedule quality. Finally, Ren and Zhang (2021) proposed a semantic rule-based approach for information extraction from procedural documents used in construction. The framework they developed, called Construction Procedural Data Integration (CPDI), was based on NLP to automate the extraction, analysis, and process the procedures prepared in text.

As this thesis is built on the FastText algorithm, a separate paragraph was allocated to this group of research using FastText. To the author' knowledge, the literature includes only three studies that used the FastText algorithm in the construction industry, according to the latest search conducted on the Scopus database on June 16th, 2022. Jallan and Ashuri (2020) adopted FastText to determine and categorize risk types in publicly traded construction firms. They utilized the FastText algorithm for transforming text into vectors that machines can understand. In addition, they used cosine similarity for classification. Li et al. (2020) also adopted the FastText algorithm in text classification to categorize safety-related documents based on their content. Their aim was to expedite the court process. However, unlike the current study, they used the supervised training version of FastText. Finally, Hong et al. (2021) focused on analyzing schedule patterns of construction projects. They stated that activity names prevent effective machine learning applications in construction schedules as activity names are text-based data. To overcome this challenge, they compared four different clustering methods for text descriptions of activity names in construction schedules. These methods were topic modeling (latent semantic analysis (LSA) and latent Dirichlet allocation (LDA)) and language models (word2vec and FastText). However, their outcome showed that topic modeling performed better than language models when activity names in construction schedules were used.

The following graph should be emphasized before finalizing the review. Unlike the previous two graphs in this section, the graph in Figure 3.3. reveals a holistic summary of studies covering both QM and KM (in green bars). Also, the blue line highlights the research incorporating AI into their perspectives. Despite an overall increase in the number of studies in the two areas mentioned above, the publications using AI are very limited, according to the numbers represented by the blue line. This thesis aims at contributing to this specific part of the literature.

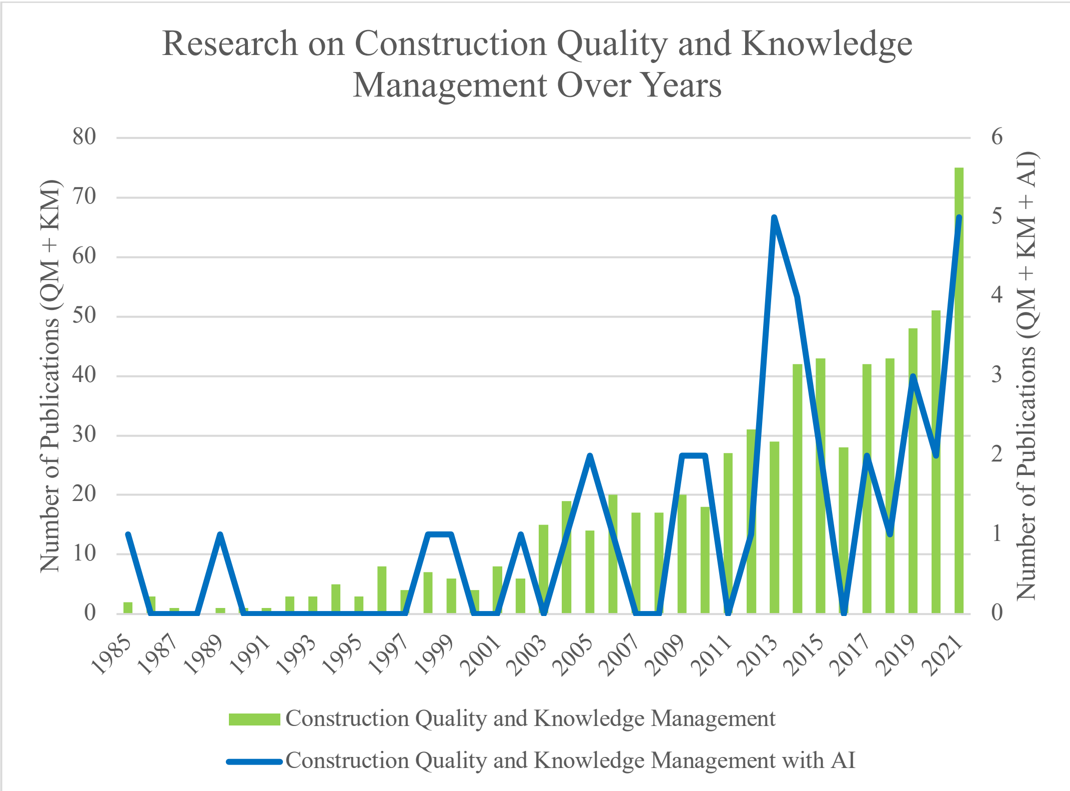

Figure 3.3. Research Trends on Construction QM and KM Between 1985 and 2021

The thesis comprises two novel approaches to the construction industry. The first one is merging different QM methods into a single strategy. Currently, NCRs, LLs, and audit findings are collected manually, ignoring the advantages of digitalized and standardized applications. This study proposes a state-of-the-art strategy for construction QM so that project quality can be increased by taking advantage of NLP-based KM. The second novelty comes from the model used to process text data. FastText algorithm has only been used in three studies (Hong et al., 2021; Jallan & Ashuri, 2020; Li et al., 2020). However, neither of them had an interest in construction QM applications. The thesis also aims to enrich the literature on NLP applications in construction QM since there are few studies in this research field. The framework and the tool developed in this thesis will provide a new perspective on QMS in the construction industry.

CHAPTER 4

RELATIONSHIP BETWEEN KNOWLEDGE AND QUALITY

The core of this thesis is the assumption that the quality of construction projects and processes can be improved with an effective and efficient KM strategy. Therefore, understanding the synergy between KM and QM and how one supports the other should be elaborated. The argument was supported by two sources: the internal audit findings conducted at the Company and the literature.

The Company's internal audit findings conducted in 2018 showed that the top three root causes leading to nonconformities were, in order, missing records, violation of a requirement or procedure, and defective or inadequate procedure. Moving onto the year 2020, the results showed that the same three causes led to nonconformities. These results are provided in Figure 4.1. and Figure 4.2. Evidently, there needs to be a system that assists decision-making while following certain procedures, and the data should be recorded when best practices or nonconformities are observed.

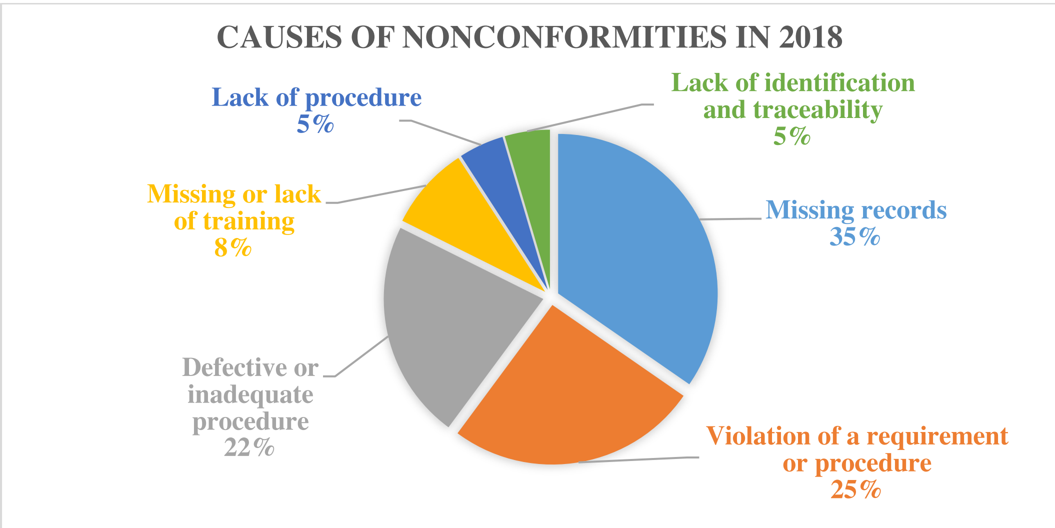

Figure 4.1. Causes of Nonconformities in 2018

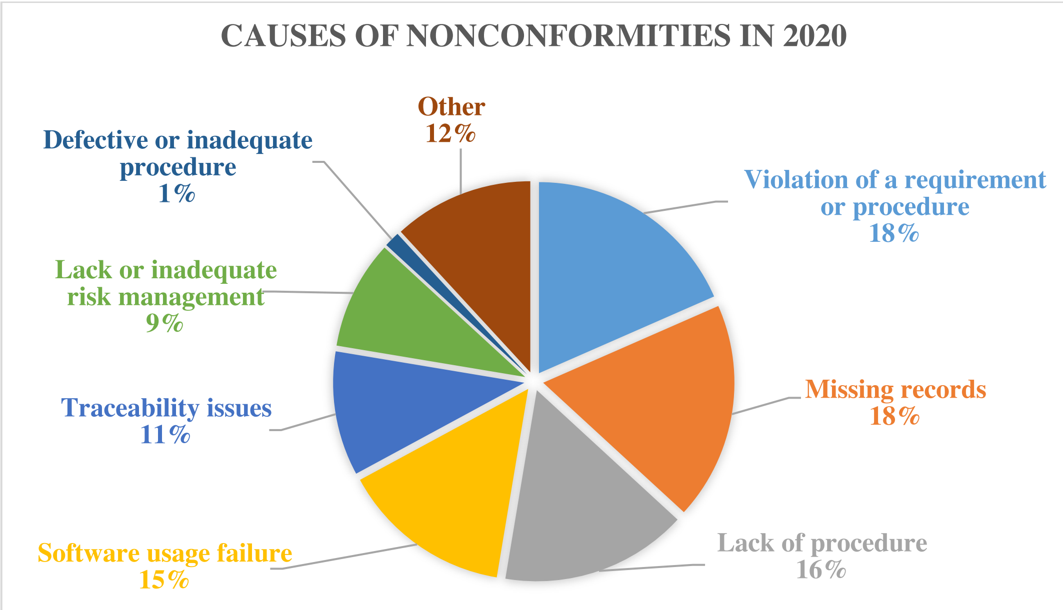

Figure 4.2. Causes of Nonconformities in 2020

In addition to the Company's findings, previous research also supports the arguments that inadequate KMS has a negative impact on the quality and that having a well-functioning KMS can improve the quality level of projects and processes. Studies showed an overlap between the contributions of KMS and factors that result in higher quality. It should be reminded that what is meant by higher quality is an increase in project performance, compliance with standards, reduced rework, reduced waste, higher stakeholder satisfaction, and other parameters that keep construction projects in line compared to their initial plan.

It was reported that KM has benefits in better decision-making and dissemination of best practices (*Knowledge Management Research Report*, 2000; Siemieniuch & Sinclair, 1999), which improve the overall quality of construction processes. The same report (*Knowledge Management Research Report*, 2000) also stated that KM improves employee skills. According to C. O. Egbu (1999), decision-making and communication are among the top five most important skills in construction. As these are the contributions the report argued to be the results of KM, how KM fosters employee skills can be understood. Also, employee skills have a direct influence on project performance and quality (Chan et al., 2004), which further supports the correlation between KM and QM.

Moreover, the fragmented structure of the construction industry leads to inefficiencies, dissatisfied customers, and low profit (Egan, 1998; Kamara et al., 2002). As a remedy, KM can enable fulfilling client needs and preserving the level of competitiveness in the industry through increased organizational performance (P. M. Carrillo et al., 2002; Hallowell, 2012; Kamara et al., 2002; Tserng & Lin, 2004). Also, with enhanced communication (Siemieniuch & Sinclair, 1999), the impact of fragmentation can be reduced.

Knowledge reuse is vital in improving the quality of solutions to construction-related problems (Tserng & Lin, 2004; Tupenaite et al., 2008). On the other hand, loss of knowledge leads to a higher risk of encountering rework and waste of effort (Siemieniuch & Sinclair, 1999). Proper KM can lower the rate of rework and nonconformity reoccurrence (Tserng & Lin, 2004).

Ultimately, it would be logical to improve the applied KMS so that the quality of companies' actions and products could be higher. Table 4.1. expands and summarizes the potential contributions of KM in the construction industry by giving reference to the related literature.

Table 4.1. Potential Contributions of KM in Construction

Contributions of Knowledge Management	**Reference**
Increasing competitive advantage	(al Qady & Kandil, 2010; C. Egbu et al., 2001; C. O. Egbu, 2004; Hallowell, 2012; Nonaka & Takeuchi, 1995; Oltra, 2005; Tupenaite et al., 2008)
Better decision-making	(*Knowledge Management Research Report*, 2000)
Improving employee skills	(*Knowledge Management Research Report*, 2000)
Dissemination of best practices	(*Knowledge Management Research Report*, 2000; Siemieniuch & Sinclair, 1999)
Producing innovation	(C. O. Egbu, 2004; Tserng & Lin, 2004)
Reducing project time	(Abdul-Rahman et al., 2008; Tserng & Lin, 2004)
Customer satisfaction	(Kamara et al., 2002)
Improved communication	(Siemieniuch & Sinclair, 1999)
Improved organizational performance	(P. M. Carrillo et al., 2002; Hallowell, 2012; Tserng & Lin, 2004)
Lower chance of nonconformity reoccurrence	(Tserng & Lin, 2004)

CHAPTER 5

METHODOLOGY

To address the point of departure of this research and answer the research questions, Cross-Industry Standard Process for Data Mining (CRISP-DM) was followed as a methodology in this thesis. CRISP-DM is commonly used in various industries and studies that revolve around data (Poh et al., 2018; Wirth & Hipp, 2000). It possesses a comprehensive and iterative approach that aims at achieving data mining goals regardless of the industry to which it is applied.

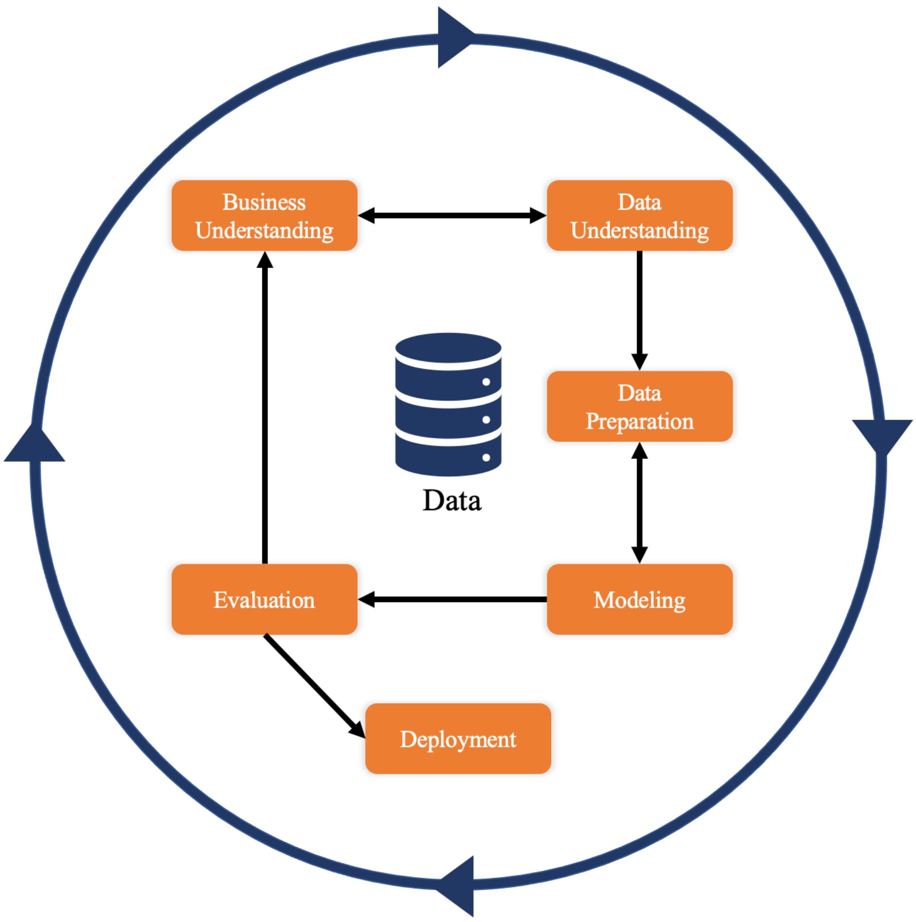

Figure 5.1. Cross-Industry Standard Process for Data Mining (CRISP-DM)

i. Business understanding

The first step of CRISP-DM is understanding the business. The methodology puts a high priority and importance on the objectives and needs of related industries. In this research, a real-world problem was defined to perform the research so that the research outcomes contribute to both the literature and the industry. Therefore, understanding how the construction industry operates was crucial.

ii. Data understanding

The second step is understanding the data. This step covers data collection, data understanding, and data exploration. To build an NLP-based recommendation tool and a knowledge base (KB), data was collected from the Company. This data included internal audit findings and lessons learned information. Code scripts were developed to understand and explore the data. In general, the missing or irrelevant data and the overall quality of text data were assessed in this step. Python language was used during the entire study due to including open-source libraries that are useful for NLP applications (Jallan & Ashuri, 2020).

iii. Data preparation

As in other NLP studies, data preparation (or preprocessing) was crucial. Unlike similar studies, the research did not require train-validation-test separation since a pre-trained language model whose validity was approved by the literature was adopted. During the preparation of data, three main paths were followed based on the nature of the data. In either case, common Python libraries were utilized (e.g., NLTK, Numpy, and Pandas).

If the data belonged to a nominal categorical type, one-hot encoding was applied. If it was in the ordinal categorical group, those variables were mapped. In the final path, if the data was in an unstructured format (text data), the data was tokenized, lowercased, and stopwords and punctuation were removed from the text to increase retrieval success. Also, detecting the language of the data entry was necessary to determine which language model to use while vectorizing the text. While detecting the language, the FastText library (lid.176.ftz) was used (Joulin, Grave, Bojanowski, & Mikolov, 2016; Joulin, Grave, Bojanowski, Douze, et al., 2016). The final preparation step was vectorization. Two language-specific pre-trained models of the FastText library were used during text vectorization. These models were pre-trained using two corpora, one for the English language (cc.en.300.bin) and the other for the Turkish language (cc.tr.300.bin) (Grave et al., 2018).

iv. Modeling

The main objective of the proposed model is information retrieval, where the similarity of various case sets should be assessed. However, a model was not needed for this purpose; instead, a function that assessed the similarity of vectors was sufficient. This study adopted cosine similarity to assess the similarity of different cases so that the most appropriate recommendation could be made. Therefore, this step of the CRISP-DM methodology was replaced with the use of a function.

v. Evaluation

The validity of the used pre-trained model was assessed by the literature; therefore, the only evaluation necessary was whether the designed approach was able to address the quality problems of the industry. To perform this assessment, two evaluation techniques were used. The first approach analyzed the overdue nonconformities, and the second approach included a survey conducted with experts from the industry. A single-question survey was prepared to observe the impact of the proposed QMS with respect to the problems of the industry. The problems covered both the factors leading to insufficient quality and the ones that were the result of poor quality. Overall, the survey assessed to what extent the study could impact its domain.

vi. Deployment

This step includes deployment, monitoring, maintenance, review, and planning of these processes. The developed tool was designed for deployment for the use of construction companies. However, its deployment was left for a future study. The deployment of Python code will be performed by Flask API.

CHAPTER 6

DEVELOPMENT PROCESS

6.1 Business Understanding

The research started with the idea that could improve QM procedures. The initial thought was that an NLP-based recommendation tool could make QM more effective and efficient in construction companies. The Company approved the validity of the research statement, stating that there is an existing need for a system that assists decision-making in handling quality issues observed. The affirmative answer indicated that the research could have contributions and practical implications. Therefore, the outcomes of this methodology step were a clear problem statement and points of departure expected to be addressed.

6.2 Data Understanding

As in most data-driven research, this study had a major dependency on the quantity and quality of data. Therefore, the research continued with determining the kind of data needed, collecting it, and understanding the data. Here, two types of data sources were used. These data sources were internal audit findings and lessons learned files. Each source had a different perspective in terms of the content of the information.

The Company continuously performs internal audits on its projects, corporate processes, and subsidiaries to identify nonconformities, best practices, and lessons learned. The Company's activities and processes are compared with its internal quality standards, ISO 9001:2015, and other regulatory bodies by which the Company abides. Project audits are performed at project sites and site offices to measure the conformance of activities performed. Similarly, corporate process audits evaluate the conformance of activities and processes of headquarters. Finally, subsidiary audits are conducted for the same purposes but with specific attention to subsidiary activities. Although the internal audit procedure followed by the Company includes more audit types, these were not within the scope of the study. Also, the scope of the thesis included only the nonconformances and excluded best practices. The collected data consisted of 1188 audit findings recorded between the years of 2018 and 2019 and 506 lessons learned cases that have been recorded since 2009. Although certain columns were missing in the data, missing values did not constitute any major problem for the ones that were used in the study. The details on this issue will be elaborated further in the following subsections.

6.3 Data Preparation

The success of the study depended on the quality of that data as much as its quantity. To achieve the highest possible data quality, preprocess was needed. Initially, unnecessary attributes were removed from the dataset. The remaining attributes were given in Table 6.1. and Table 6.2. The ‘Attribute’ columns list the attributes used. The ‘Data Type’ columns indicate whether an attribute is nominal categorical, ordinal categorical, or text. Finally, the right-most columns show the structured nature of attributes. The identification of type was crucial while preprocessing.

Table 6.1. Audit Finding Attributes

Attribute	Data Type	Data Type Category
Business Line	Nominal Categorical	Structured
Related Department	Nominal Categorical	Structured
Finding Type	Ordinal Categorical	Structured
Finding	Text	Unstructured
Root Cause	Text	Unstructured
Correction	Text	Unstructured

- Business Line: It shows the project group (whether it is a power project, oil and gas project, civil work, or an infrastructure project) where a given nonconformity was detected. It may also show that a nonconformity was in office functions (represented as Headquarters),
- Related Department: It shows the department where the finding was detected,
- Finding Type: It indicates whether the finding is major (MJR), minor (MNR), or classified as an observation (OBS),
- Finding: It describes the finding by explaining the nonconformity,
- Root Cause: It shows the underlying root cause leading to a given finding,
- Correction: It explains the action taken to correct the identified nonconformity.

Table 6.2. Lessons Learned Attributes

Attribute	Data Type	Data Type Category
Business Line	Nominal Categorical	Structured
Related Department	Nominal Categorical	Structured
Lessons Learned	Text	Unstructured
Root Cause	Text	Unstructured

- Business Line: It shows the project group (whether it is a power project, oil and gas project, civil work, or an infrastructure project) where a given lesson was learned. It may also show that a lesson was learned from office functions (represented as Headquarters),
- Related Department: It shows the department where a given lesson was recorded,
- Lessons Learned: It explains the learned lesson,
- Root Cause: It shows the root cause of a given case that leads to a learned lesson.

After the dataset was narrowed down to the given attributes, preprocess continued based on the data type. Different types of data require different techniques while preprocessing so that the entire data set can be processed uniformly. These different data types can be divided into two major groups: structured and unstructured data, and the methods used while preprocessing step are investigated in these two sections. It should be noted that some of the records did not include root causes (11% missing) or correction methods (%15 missing); however, the missing information did not cause any significant issues for the study. The missing records only affected the richness of the KB and the relevance of retrieval.

6.3.1 Handling Structured Data

The structured data used in this thesis incorporated nominal categorical (Business Line and Related Department) and ordinal categorical data (Finding Type). The difference between the two is that nominal categorical data does not have an internal structure; in other words, one category is not larger or more important than the others. However, ordinal categorical data requires an order since one category might be more significant than the other. Due to this difference, two separate preprocessing techniques were applied in this study.

For the ordinal categorical data, the order between categories was taken into account. Therefore, in this study, the lowest level category was assigned a "1", and the highest level category was assigned the size of the category set in that attribute. In this case, the following mapping was applied to the Finding Type column:

$$\{OBS \rightarrow 1; MNR \rightarrow 2; MJR \rightarrow 3\}$$

For categorical data, dummy coding (one-hot encoding) was applied. This method can be explained clearly with an example. Assuming there are three categories in a dataset, namely A, B, and C, dummy coding generates three new attributes whose names are the corresponding category names. Then, if the category of an instance is A, the method assigns "1" to column A and "0" to the others. Similarly, if the category of an instance is B, dummy coding assigns "1" to column B and "0" to columns A and C. The following indicator function represents the method. Here, $\mathbb{1}_A : X \rightarrow \{0,1\}$.

$$\mathbb{1}_A(x) := \begin{cases} 1 \; if \; x \in A, \\ 0 \; if \; x \; \notin A. \end{cases}$$

In this study, the collected nonconformity cases were grouped under six business lines, namely, ‘power’, ‘infrastructure’, ‘oil, gas, and petrochemicals’, ‘civil’, ‘headquarters’, and ‘subsidiaries’. The first four categories show the business line of the project in which the audit had been conducted. The final two categories indicate that the finding belongs to either headquarters or its subsidiaries. In the light of this information, six dummy variables were created for the ‘Business Line’ column. In addition, the information under the ‘Related Department’ column was categorized under 31 groups; therefore, there were 31 unique variables in the ‘Related Department’ column. As a result, 31 additional dummy columns were added to both datasets. With the addition of the mapped ‘Finding Type’ column, there were 38 variables in the audit finding dataset that will be used for similarity assessment in the following steps. Lessons learned dataset was transformed into the exact same number of features except for ‘Finding Type’, which did not exist.

Table 6.3. Feature Breakdown for Structured Data

Feature Name	Number of Columns (Audit Findings)	Number of Columns (Lessons Learned)
Finding Type	1	0
Business Line	6	6
Related Department	31	31

6.3.2 Handling Unstructured Data

The unstructured data may exist in image, sound, or text files. The reason why it is called ‘unstructured’ is that it lacks the structure that computerized methods can process (Gandomi & Haider, 2015), which makes it necessary to transform this type of data into a structured format. Therefore, this section was one of the most crucial parts of this research since the entire model relied on text processing.

There are common text preprocessing steps that were used in most of the NLP studies (Caldas et al., 2002; Candaş & Tokdemir, 2022a; Jallan & Ashuri, 2020). In this study, (i) tokenization, (ii) lowercasing, (iii) removing stopwords, (iv) removing punctuation, double spaces, and empty strings, and (v) vectorization were applied to the text data in both datasets. Also, the language of the text in each instance was detected throughout the preprocess.

The first step of preprocessing text was tokenization so that each word can be preprocessed. Tokenization was done by using the NLTK library. After lowercasing each word, stop words were removed as these words are usually abundant and affect the quality of results negatively (Shin & Issa, 2021). However, that the datasets included text data with both English and Turkish sentences challenged having accurate results. For this reason, a language detection module was inserted into the preprocessing phase. Based on the detected language, stop words belonging to the given language were removed from the text. In addition, punctuation was also removed since it was not necessary for similarity assessment (Candaş & Tokdemir, 2022a). Tokens with empty strings were removed to cleanse the final datasets. The following diagram summarizes the entire preprocessing pipeline that was used in this research.

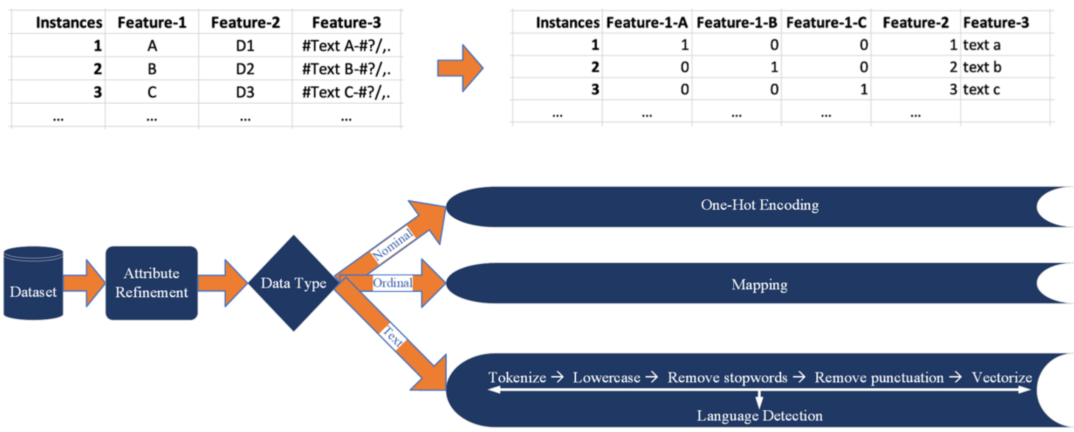

Figure 6.1. Data Preprocess Pipeline

6.3.3 Vectorization (Feature Extraction)

The final step of data preparation was vectorization. This step was also called feature extraction since text-based data was transformed into a vector, where each value in the vector represents certain information from the original text. Therefore, this step was also among the most important processes of this study. Text-based information in audit findings and lessons learned files was collected in two separate datasets and given to the FastText model to be vectorized. For text in Turkish, the model "cc.tr.300.bin" was used, and for text in English, "cc.en.300.bin" was used (*FastText*, 2020). The algorithm of the original model was given below according to the explanations of Bojanowski et al. (2016).

Considering a word group of size W, where w represents each word (i.e., $w \in \{1,2,\ldots,W\}$), the ultimate objective of the Skip-Gram model is to vectorize w by maximizing the following:

$$\sum_{t=1}^{T} \sum_{c \in C_t} \log p(w_c | w_t) \qquad (1)$$

Here, C_t represents the word set that are predecessors and successors of the word w_t, and the equation calculates the probability of observing w_c, given the word w_t. To develop a relationship between context and words, a scoring function, s, is used. The softmax function is used for the mentioned probability.

$$p(w_c | w_t) = \frac{e^{s(w_t, w_c)}}{\sum_{j=1}^{W} e^{s(w_t, j)}} \qquad (2)$$

The above equation assumes that there is only one context word corresponding to a word, w_t. Therefore, the binary logistic loss function is used to individually classify whether a context word exists given a word, w_t.

$$\log\left(1 + e^{-s(w_t, w_c)}\right) + \sum_{n \in \mathcal{N}_{t,c}} \log(1 + e^{s(w_t, n)}) \qquad (3)$$

In Equation 3, $\mathcal{N}_{t,c}$ denotes the set of context words that are not existing given a word w_t. These words are randomly selected from the vocabulary. The Equation-3 can be rewritten with the logistic loss function, $l(x \rightarrow \log(1 + e^{-x}))$.

$$\sum_{t=1}^{T}\left[\sum_{c\in C_t} l(s(w_t, w_c)) + \sum_{n\in\mathcal{N}_{t,c}} l(-s(w_t, n))\right] \tag{4}$$

Word vectors are used in the scoring function, s. Assuming two vectors, u_{w_t} for w_t and v_{w_c} for w_c, the scoring function between w_t and w_c can be written as the scalar product of the two vectors.

$$s(w_t, w_c) = u_{w_t}^{\mathsf{T}} v_{w_c} \tag{5}$$

The above-mentioned algorithm belongs to the Skip-Gram model (Mikolov, Sutskever, et al., 2013). However, Bojanowski et al. (2016) argued that the skip-gram model ignored the words' internal structure and proposed another scoring function. In their model, they used the bag-of-character n-gram model to represent each word in the vocabulary. While doing that, they inserted predefined characters to separate each word from its neighbors. They also added the word itself to its n-grams. The entire n-gram concept can be thought of as if a word, w, is the set and its n-grams are its subset in addition to the word, w, itself. The following example can clarify the topic. Given a word <north> and n=3, the n-grams are as follows. These subsets are what represent the word <north>.

<north> → <no, nor, ort, rth, th> + <north>

According to the modification Bojanowski et al. (2016) made, a new dictionary is assumed with size G. For a word, w, where $\mathcal{G}_w \subset \{1,2,\ldots,G\}$) is the set of n-grams, a new vector, z_g is created for each n-gram, g. Therefore, each word is represented by the sum of its n-grams' vector representations.

$$s(w, c) = \sum_{g\in\mathcal{G}_w} z_g^{\mathsf{T}} v_c \tag{6}$$

As a final computational detail, the developers of the FastText model assigned each n-gram to an integer. The result was that each word was represented by its location in the sentence, called index, and its integer set. This study used a model that was pre-trained using this algorithm to vectorize text in the dataset. Therefore, additional training was not necessary for the research.

It was proven that more accurate similarity-based retrieval was possible when a FastText model was used (Bojanowski et al., 2016). This is also valid for technical and rare words. In addition, the underlying algorithm of the FastText model performs subword analysis, meaning that it searches for not only the words but also word chunks so that a meaningful interpretation can be made (Bojanowski et al., 2016). Therefore, the FastText models were applied to both datasets. The 'Findings', 'Root Causes', and 'Corrections' columns in the audit findings dataset and the 'Lessons Learned' and 'Root Causes' columns in the lessons learned dataset were converted into separate 300-column vectors.

At the end of this phase, the entire KB was transformed into a structured format. Audit findings were represented by (i) a 38-column and 1188-row dataset for 'Finding Type', 'Business Line', and 'Related Department', (ii) a 300-column and 1188-row dataset for 'Findings', (iii) a 300-column and 1188-row dataset for 'Root Causes', and finally (iv) a 300-column and 1188-row dataset for 'Corrections'. Similarly, lessons learned data was represented by (i) a 37-column and 506-row dataset for 'Business Line', and 'Related Department', (ii) a 300-column and 506-row dataset for 'Lessons Learned', and finally (iii) a 300-column and 506-row dataset for 'Root Causes'.

6.4 Modeling

Similarity scoring is widely used in research where the primary interest is in information retrieval. It represents the distance between the two points in an n-Dimensional space, i.e., a current case given as input and one of the previous cases in the case base (Goh & Chua, 2009). This study adopted cosine similarity to compare a new case with the previous ones and retrieve the most similar one. The reason why this particular technique was selected is that cosine similarity has proven itself in text-based similarity applications (Jallan & Ashuri, 2020; Singhal, 2001)

Cosine similarity measured the angle between two vectors as an indicator of their similarity. It was derived from the Euclidean dot product of two non-zero vectors. In this case, the vectors were text vectors generated by the FastText model. If the result was -1 (corresponds to 180˚), the vectors were in the opposite direction. On the contrary, if the cosine function outputted 1 (corresponds to 0˚), it could be understood that the vectors were the same. The following figure helps understand the working mechanism of the function in a visual. However, it should be reminded that the vectors used in this study had 300 dimensions. Cosine similarity was also used with FastText models and proved its accuracy (Bojanowski et al., 2016). Therefore, there was a fit between the extraction and retrieval methods used in this study.

$$A \cdot B = \|A\| x \|B\| x \cos\theta \tag{7}$$

$$\cos\theta = \frac{A \cdot B}{\|A\| x \|B\|} \tag{8}$$

$$\cos\theta = \frac{\sum_{i=1}^{n} A_i \cdot B_i}{\sqrt{\sum_{i=1}^{n} {A_i}^2} \cdot \sqrt{\sum_{i=1}^{n} {B_i}^2}} \tag{9}$$

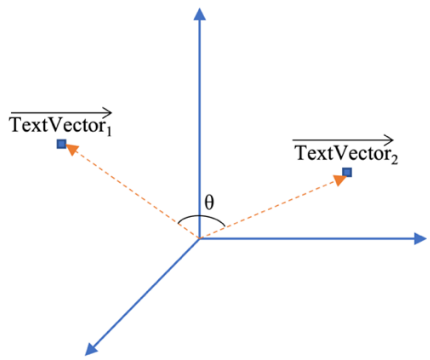

Figure 6.2. Cosine Similarity of Two Text Vectors in A Three-Dimensional Space

This step concluded the overall mechanism of the proposed algorithm. With the help of the pre-trained vectorization model and the similarity function, this study was built on similarity-based information retrieval in recommending corrections and preventive actions to users. When a new case was inputted, the input data was processed with the preprocessing pipeline that was explained in the last part. Depending on the use case, the tool provided the most similar n cases, where n represents the number of recommendations that the user prefers.

CHAPTER 7

INTEGRATION AND IMPLEMENTATION

ISO 9000 guidelines dictate that QMS should be integrated into the existing strategies of organizations. Since this study proposes a tool to support and assist decision-making in quality practices rather than altering the entire business strategy, integration was also an essential part of the study. To achieve a successful integration, both the designed framework and the existing system should be well-understood. The technical details of the proposed tool were already explained. Therefore, there remains a need for a system-in-place that is used for QM purposes. In this thesis study, the QM strategies of the Company were used to propose an overall strategy for KM-enabled QM. Although the design seems specific to the selected company, it can be implemented by any company following a standardized approach.

7.1 System Design

The tool can be adopted in two parts of QM practices, which results in an overall QM strategy for companies. The first part aims at preventing nonconformities by making proactive suggestions to users. In this module, the system considers what is ahead by reading the inspection requirements. Therefore, this part can be thought of as a QA practice. The second part plays a corrective role during QC. Here, the aim is to handle nonconforming activities or processes after they occur. Based on the previous actions taken to correct nonconformities, the tool recommends possible solutions to a given nonconformity. Overall, a well-designed combination of the two consecutive parts creates a continuous learning and continuous improvement cycle in which nonconformities can be minimized. Ultimately, the integrated design aims at a state of Poka-Yoke, i.e., mistake-proofing in construction projects. The following paragraphs explain the details of the two parts of the proposed QMS.

7.1.1 Part-1: Prevention Mechanism

In the first part, the framework is interested in making suggestions before a selected work item is executed to prevent potential nonconformities and resultant rework and waste. The module can apply to both the project side and corporate side. When the project side is considered, the process starts in the early phases of construction projects with the definition of inspection and test plans (ITP). ITPs include certain requirements to be checked on site. These requirements are stated in text format, which the NLP-based model can analyze. With this tool, a given requirement sentence is preprocessed and vectorized so that the words are converted into numbers, to which mathematical operations can be performed. Then, the cosine similarity of the given sentence vector is calculated with respect to previous cases in the KB, and the one with the highest similarity score is retrieved. As a result, the lessons learned in the KB are used as preventive actions before the work is executed. The process is almost the same for the corporate side. The only difference is the input. In this case, audit checklists are used as inspection requirements. The output is, again, retrieved from the KB.

There are also additional inputs so that the retrieval tool understands whether it is a corporate-related process or a construction activity. These additional inputs are “Business Line” and “Department of Process/Activity”. These inputs help make more relevant suggestions to a given case. Depending on preference, the number of actions suggested to users can be altered. The next item in the flow is ensuring that applicable recommendations are satisfied before the activity or process is executed. This part is crucial since, without implementation, the suggestions are worthless. After QC, if the process or activity under investigation conforms to quality standards, the module repeats for the next process/activity. On the other hand, when there is nonconformity, the second module starts.

7.1.2 Part-2: Correction Mechanism

The second part focuses on corrections to be applied in case a nonconformity is observed during or after the execution of a selected work item. When an activity or process is found not to be compliant, the findings are recorded into NCRs and audit findings lists digitally. Here, in addition to the explanations of a detected nonconformity, the type of finding and its root cause are also recorded. This information is combined with the business line of the process/activity and the related department and stored in the KB. Simultaneously, the input is preprocessed, the text is vectorized, and the cosine similarity function is called. The function returns the most similar case based on the cosine of the two vectors.

Here, there are two possible scenarios. The user may or may not find the recommended action suitable to the existing nonconformity. If the suggestion is not found valid, the user can provide manual input. On the contrary, if the tool makes a proper recommendation, the user can implement the action immediately. In either scenario, the new instance is appended to the KB. The process goes back to the QC, where the conformity of the work is assessed, after which the same cycle applies.

The two modules are combined in Figure 7.1. The straight lines represent data transfer from one process to another, and the dotted lines indicate data transfer from or to the KB. The orange blocks represent the steps where NLP-based information retrieval is used in the overall design. The flow chart that includes technical details is given in Appendix-A. The KB structure and the input-output flow are shown in Figure 7.2 and Figure 7.3, respectively. In Figure 7.3, yellow blocks represent the input given to the recommendation tool, blue blocks show the outputs of the tool, and yellow one shows the KB feeding the tool. The straight line represents the data flow within the prevention module, and the dashed line shows the data flow for the correction module. Overall, the designed QMS improves knowledge transfer within the company and leads to continuous improvement and continuous learning cycle in the work environment (see Figure 7.4.).

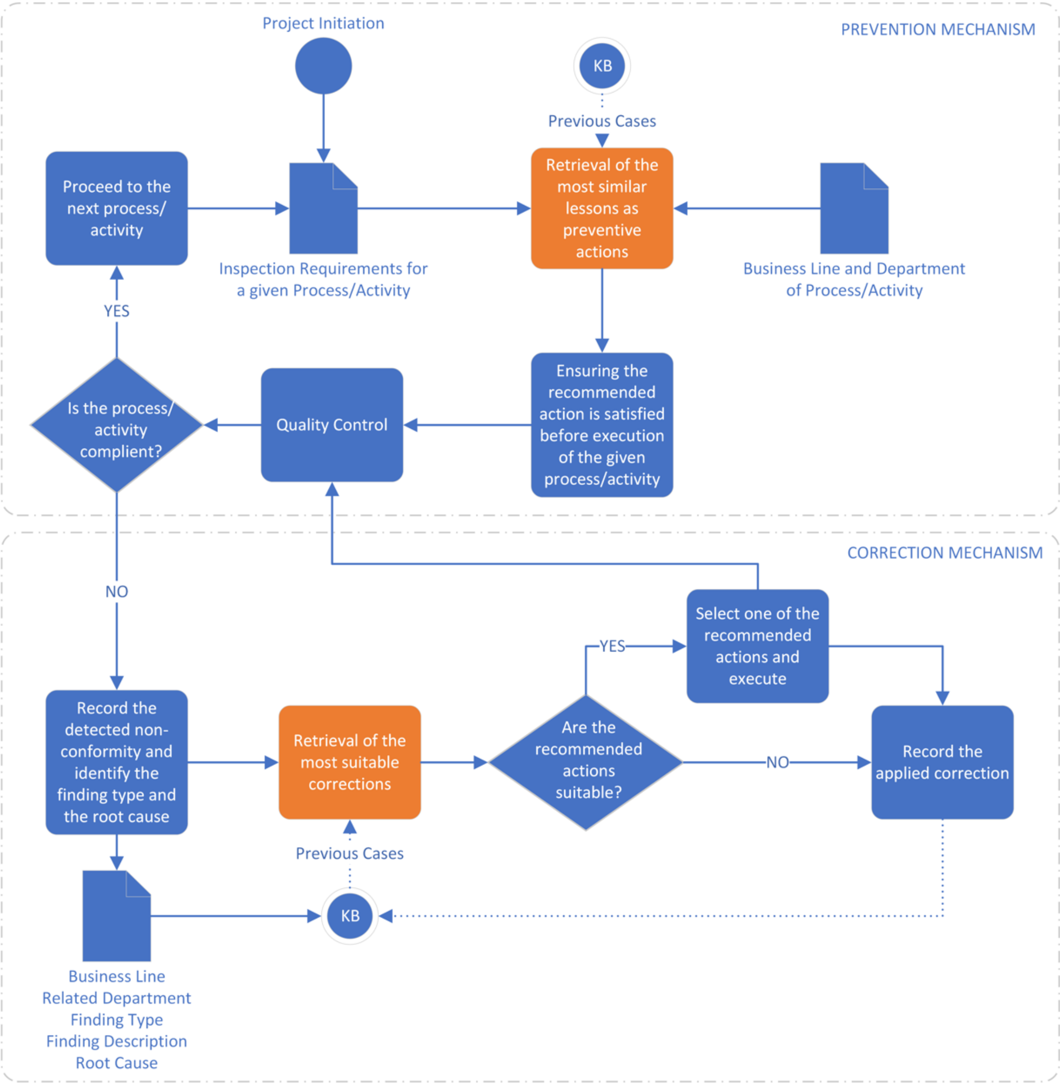

Figure 7.1. Integration into Overall Quality Management Strategy

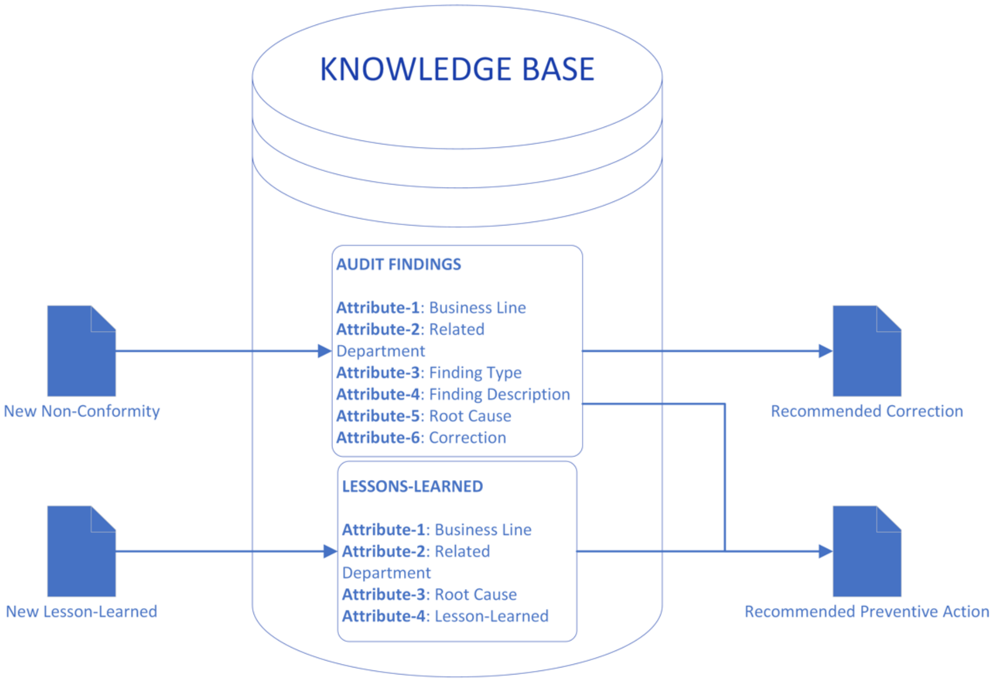

Figure 7.2. The Knowledge Base (KB) Structure

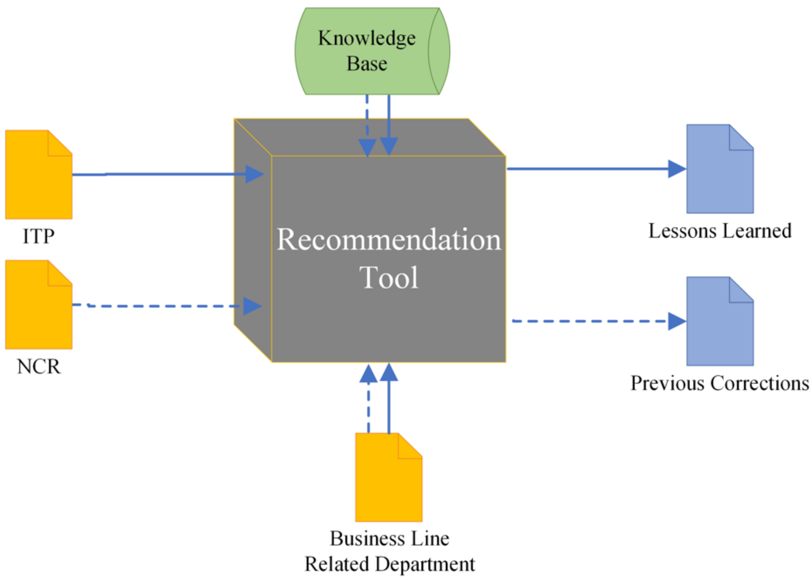

Figure 7.3. Input-Output Flow of the Proposed System

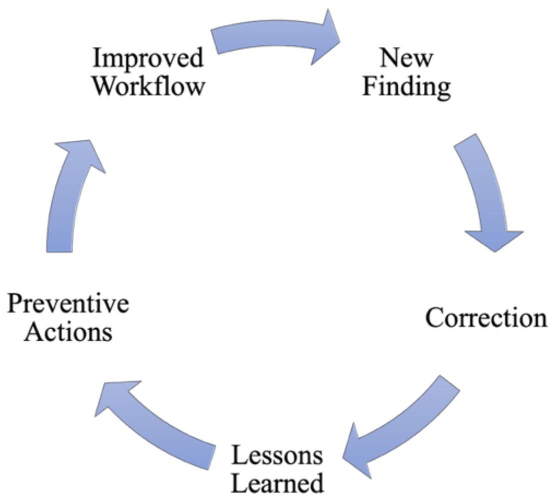

Figure 7.4. Continuous Improvement Through Learning

The chapter continues with three scenarios that exemplify and clarify the use cases of the tool. The first scenario shows how the tool suggests proactive measures to prevent a potential nonconformity. This scenario relates to continuous improvement. The second scenario exhibits the tool's ability to provide to-the-point corrections when nonconformity is observed. Here, it was shown that the tool is error-proof when human-born mistakes during information recording are considered. A real nonconformity case is tested, and the suggested action was compared with the actual correction applied to the case. Finally, the third scenario explains how the tool can be used in the case the recommendations are inapplicable to the problem at hand. This scenario also shows the dynamic retrieval mechanism of the tool and clarifies its impact during information recording and continuous learning. It should be noted that the three scenarios reflect the significant features of the tool; however, there might be additional scenarios depending on the use case.

7.2 Implementation Scenario – 1: Prevention

In this scenario, the following inspection requirement was taken as an example case. This case was taken from the ITP for the fence installation process that was applied in an oil and gas project.

Inspection Requirement:

"Issue inspection request for the final walk-down to inspect the pump in order to form (official) final punch list."

Firstly, the text was vectorized into the following structure (Table 7.1.). Each value represented a feature extracted from the text. As explained previously, there were 300 features by default when a FastText model was used. These values were replicable, that means, when the same pre-trained model is used by another researcher for the given text, the same values can be observed. However, it should be noted that the text should be preprocessed in the same method; otherwise, the words and the characters of the text might differ, so would the text vector.

Table 7.1. Unstructured Input for Scenario-1

Text					
Inspection Requirement					
"Issue inspection request for the final walk-down to inspect the pump in order to form (official) final punch list."					
Feature-1	**Feature-2**	**Feature-3**	**…**	**Feature-299**	**Feature-300**
-2.483 e-03	3.539 e-02	2.466 e-02	…	-1.740 e-02	9.123 e-03

In addition to the inspection requirements, the business line and the related department of the activity were given as additional input. These values were nominal categorical variables for the information retrieval algorithm.

Table 7.2. Structured Input for Scenario-1

Categorical	
Business Line	**Related Department**
Oil and Gas	Quality

After the similarity function assessed the cosine similarity of the input and the previous cases, the following output is given. It should be noted that the number of recommended actions can be increased depending on the preference.

Table 7.3. Output For Scenario-1

Output	[-9.850 e-03, 2.592 e-02, … ,-1.764 e-02, 1.173 e-04]
Similarity Score	0.710
Index	31
Text	No punch lists are prepared until the end of system completion. Too many punch items are left to acceptance stage. As no walkdowns are performed and no punch lists are prepared during the construction period, we end up with a lot of punch items which become evident during the turnover and handover stage of the systems. This effects system handover dates.

Once the suggested action is taken, if applicable, the rest of the procedure consists of existing QC practices. This scenario shows the proposed model's ability to suggest proactive measures so that nonconformity or rework does not harm the project progress.

7.3 Implementation Scenario – 2: Correction

In this scenario, it is assumed that the first part could not help prevent a nonconformity, and the following case (Table 7.4.) was observed. This case was taken from the audit finding list that also fed this study. Although the tool searched for a solution from the same KB that the test case exists, this test case was left in the KB intentionally to observe whether the tool functions properly and returns the test case itself as the most similar solution. The structured and unstructured parts of the input in Table 7.4. were given in Tables 7.5., 7.6., and 7.7.

Table 7.4. Nonconformity Case Feeding Scenario-2

Attribute	**Data**
Finding	During site tour at level 1, a site personnel were observed using a grinder while he was sitting on a belt that was tied on re-bars, instead of using a mobile scaffold. Also at the same area, a foreman was observed using a grinder without hand and face protection, while he was not designated to use hand tool.
Root Cause	Defective or inadequate procedure
Finding Type	Major
Related Department	Construction Management
Business Line	Civil

Table 7.5. Unstructured Input-1 for Scenario-2

Text-1					
Finding					
"During site tour at level 1, a site personnel were observed using a grinder while he was sitting on a belt that was tied on re-bars, instead of using a mobile scaffold. Also at the same area, a foreman was observed using a grinder without hand and face protection, while he was not designated to use hand tool."					
Feature-1	**Feature-2**	**Feature-3**	**…**	**Feature-299**	**Feature-300**
7.983 e-03	3.502 e-02	4.127 e-03	…	5.969 e-03	3.717 e-03

Table 7.6. Unstructured Input-2 for Scenario-2

Text-2					
Root Cause					
"Defective or inadequate procedure"					
Feature-1	**Feature-2**	**Feature-3**	**…**	**Feature-299**	**Feature-300**
-4.817 e-02	-2.919 e-03	-6.674 e-03	…	1.169 e-02	4.056 e-02

Table 7.7. Structured Input for Scenario-2

Categorical		
Finding Type	**Related Department**	**Business Line**
Major	Construction Management	Civil

Each group was assessed with its peers. That means the 300-column text vector extracted from "finding" data was compared with the other finding vectors. Similarly, the root cause vector was compared with previous root causes, and finally, categorical variables were compared with other categorical vectors. As a result, the mean of three similarity scores was calculated, and the case with the highest overall score was given as output. With the above input, the following case was the most similar one.

Table 7.8. Output Similarities for Scenario-2

Output		
Text-1 Similarity	**Text-2 Similarity**	**Categorical Similarity**
0.799	1.000	0.870

$$\textit{Overall Similarity} = (0.799 + 1.000 + 0.870)/3 = 0.890$$

As a result, the outputted case had 89% similarity when it was compared with the input case. The output case and the corresponding correction were given below.

Table 7.9. The Most Similar Output for Scenario-2

Attribute	**Data**
Finding	During site tour at level 1, mobile scaffolds with no safety tags on were observed to be in use.
Root Cause	Defective or inadequate procedure
Finding Type	Observation
Related Department	Construction Management
Business Line	Civil
Correction	People were trained for scaffold usage and nine employees were trained for scaffold inspection.

As the above table shows, the output of the model was a correction method that might be a remedy for the observed nonconformity. The actual correction applied in the given project was the following:

<u>Actual Correction:</u> *'Supervisor should instruct all his personnel on duties, hazards, and how to overcome those hazards during morning Pre-Task meetings. All personnel working at 1.8m are to have fall protection training. ....”*

It was obvious that both the actual and retrieved correction methods emphasize the training of personnel. This result showed the expected use case of this tool. A final explanation concludes this scenario by showing the tool’s ability to process misspelled words. The NLP functionality of the tool enables reaching the same results with misspelled words. This feature absorbs the human error while recording nonconformity cases, lessons learned, and others. The previous case used in this scenario is altered into the following. The misspelled words were underlined.

Misspelled Input: *"During site tour at lvl 1, a site personel were observed using a grinder while he was siting on a belt that …"*

In this case, the tool gave the same output (Table 7.9.) with a slight change in similarity score, which is 0.889.

7.4 Implementation Scenario – 3: Continuous Learning

In the final scenario, the situation was a case where recommended corrections were not related to a given nonconformity, and the user wanted to insert manual input. This scenario shows the tool's ability to learn from the user input continuously. Assuming a user did not find the recommended corrections useful, they can insert a manual entry into the system. The case feeding this scenario was taken from a real case observed in a power plant project. The observed finding was as follows.

Table 7.10. Nonconformity Case Feeding Scenario-3

Attribute	Data
Finding	All joints are not acceptable visually due to the fact that joints have excessive undercut, insufficient throat, and insufficient leg. In addition, all welded joints were grinded vertically and horizontally, and sharp edges occurred. These sharp edges may cause some cracks in the future.
Root Cause	Insufficient factory inspections
Finding Type	-
Related Department	Mechanical Construction
Business Line	Power

The preprocessing steps and retrieval results were not provided to prevent repetition. In this scenario, the essential point was that the user could insert the following correction manually. When it was inserted, the case was updated in the KB, and in the future, if a similar nonconformity asks for a solution, this manual input might be recommended as well as the ones already in the KB.

Manual Input: *"Caps need to be grinded and reweld according to material specifications."*

This scenario also showed the dynamic search mechanism of the designed system. Although the "Finding Type" attribute was empty, the algorithm was able to retrieve the most similar cases. As explained earlier, when the entire attributes were available, they were given equal weights. In case one of them was missing, its weight was assigned a zero so that an empty string did not affect the search result. Although the dynamic mechanism takes missing values into account, and the results may be unaffected by this issue, it should be reminded that each data field asked to be filled is important in data-driven methodologies. Unless users record the cases in full, missing information eventually affects the search results.

CHAPTER 8

DISCUSSION AND EVALUATION

8.1 Discussion of Scenarios

The scenarios showed that the proposed tool can address multiple points of the existing QMS of construction companies. The first scenario referred to a case where the tool helped QM practitioners to take proactive action that could prevent potential nonconformities. This is one of the primary objectives of this research since the state of Poka-Yoke is aimed with a continuous improvement culture. In this ideal state, the KB is so rich that every possible lesson exists, and users can be fed with relevant and to-the-point proactive actions. Surely, this state requires a well-managed KM and digital storage of this knowledge.

The second scenario explained a case where a nonconformity had already occurred. In this scenario, the tool made certain recommendations, and one of them was selected. This shows the tool's ability to provide instant solutions to given problems. However, similar to the previous scenario, this ability necessitates an enormous KB where every possible correction is available. Again, this ability relies on having a functional KM and digitalized processes. This second scenario referred to another objective of this study: minimizing the effort while making corrections to nonconforming work items. The process of passing documents and signatures between personnel might hinder timely corrections. This challenge can be overcome with the instant solutions that the tool provides. Lastly, the example showed how the tool absorbs human-related errors while recording information. The fact that both correctly-spelled and misspelled searches retrieved the same solution showed how the tool performs better than traditional retrieval algorithms.

The third scenario showed the dynamic nature of the tool. The ideal state that the two previous scenarios relied on was not necessary here since this possibility already explained a case where recommended solutions did not apply to the given nonconformity. This was the most expected scenario since the current state of KB is either limited or non-existing. The tool's suggestions might not be related to given problems since similar problems were not encountered previously. In this growth phase of the proposed QMS, practitioners are expected to do nothing more than what they have already been doing, which is recording nonconformities, corrections, and lessons learned. However, this process should be entirely moved to a digital platform and a standardized approach so that the future can benefit from the problems of the current state. Once the KB becomes rich enough, this scenario will become faint; instead, the second scenario can become more frequent. In the long run, the second scenario will also become less frequent and leave its place to the first scenario. Ultimately, a zero-mistake state can be achieved where corrections for nonconformities are replaced by lessons learned as preventive actions.

8.2 Evaluation

The designed tool and the strategy were evaluated under two perspectives: duration-wise and impact-wise. In the first part, a duration-wise comparison was made between the current state of nonconformance correction and the proposed approach. In the second part, the impact of the proposed strategy was evaluated by experts.

8.2.1 Duration-wise Evaluation

The data in Table 8.1. belongs to the number of nonconformities observed by the Company in 2021. It was seen that there were 503 nonconforming items in 2021, of which 371 were minor and 132 were major.

Table 8.1. Number of Nonconformities in 2021

Overdue	Minor	Major	Total
No	120	39	159
Yes	251	93	344
Total	371	132	503

When minor nonconformities were considered, their overdue percentage was around 68%. Similarly, the overdue percentage of the major nonconformities observed in 2021 was around 70%. Overall, there was an approximate 68% ratio when all cases were considered. This ratio was calculated by dividing the cases that were closed after their due date by the total number of cases belonging to that category. Furthermore, when the cases whose deadlines were missed were considered, it was seen that the average overdue time is approximately 40 days for major nonconformities and 32 days for minor nonconformities. When the impact of the time taken for decision-making is considered, one of the major contributions of this study, providing instant solutions, becomes obvious.

8.2.2 Impact-wise Evaluation

The impact of the tool was evaluated by professionals in the construction industry. A single-question survey was prepared based on the factors upon which the research was built. The question asked was how applicable the tool is when the quality problems observed in construction are considered. The selected experts were given a set of factors that led to or were raised by insufficient quality. The experts were selected from the top-level managers of an ENR-250 construction company. These participants were asked to rate the impact of the tool on each problem on a 1-10 scale. The following paragraphs explain the qualifications of these experts.

Expert-1: A Vice President responsible for the quality, health, safety & environment, communications, sustainability, compliance, and information technologies. The expert has almost 23 years of executive experience and academic background with a master's degree in civil and environmental engineering.

Expert-2: A Corporate Quality Manager with a master's degree in design, engineering, and technology management and nearly 15 years of professional background. The expert is specialized in quality assurance, quality control, and auditing. The expert is also licensed in information security, quality management, and audit.

Expert-3: A Senior Information Management Supervisor with 16 years of professional experience in quality, document management, and information management. The expert holds an ISO 9001:2015 Quality Management Systems Requirements Certificate.

The result of the survey was given in Table 8.2. The average values were calculated in the right-most column, and the overall scores and the overall average was given in the bottom row. In addition, the tool's impact based on the overall score from the experts is visualized using the bar chart given in Figure 8.1. The bar chart reveals the results in descending order of overall scores.

According to the results, the challenges on which the proposed strategy can have the highest impact are:

(i) For the factors that lead to insufficient quality; 'lack of evidence-based decision making', 'lack of continuous improvement and knowledge reuse', and 'missing or inconsistent data recording',

(ii) For the factors resultant of insufficient quality; 'process delay due to poor quality', 'cost impact of poor quality', and 'rework and waste'.

The survey results also showed that the 'lack of top-management commitment' is a factor on which the proposed approach can have the least impact. Overall, the study is expected to impact the construction industry with an 8.0 score out of 10.0. This result is promising in terms of what was initially aimed.

Table 8.2. Survey Results

	Factor	E1	E2	E3	Avg.
CAUSES	Lack of continuous improvement and knowledge reuse	8	8	10	8.7
	Missing and inconsistent data recording	7	9	9	8.3
	Lack of evidence-based decision-making	8	10	10	9.3
	Lack of top-management commitment	5	5	8	6.0
	Absence of feedback mechanism	6	8	10	8.0
	Violation of requirements and procedures	6	6	10	7.3
	Defective or inadequate procedure	6	6	10	7.3
	Lack of training	6	6	8	6.7
	Lack of communication	7	6	10	7.7
EFFECTS	Process delay due to poor quality	7	10	10	9.0
	Cost impact of poor quality	7	10	10	9.0
	Rework and waste	7	10	10	9.0
	Safety issues due to inadequate quality	7	9	10	8.7
	Stakeholder dissatisfaction	5	10	9	8.0
	Loss of credibility	5	7	9	7.0
	Overall	6.5	8.0	9.5	8.0

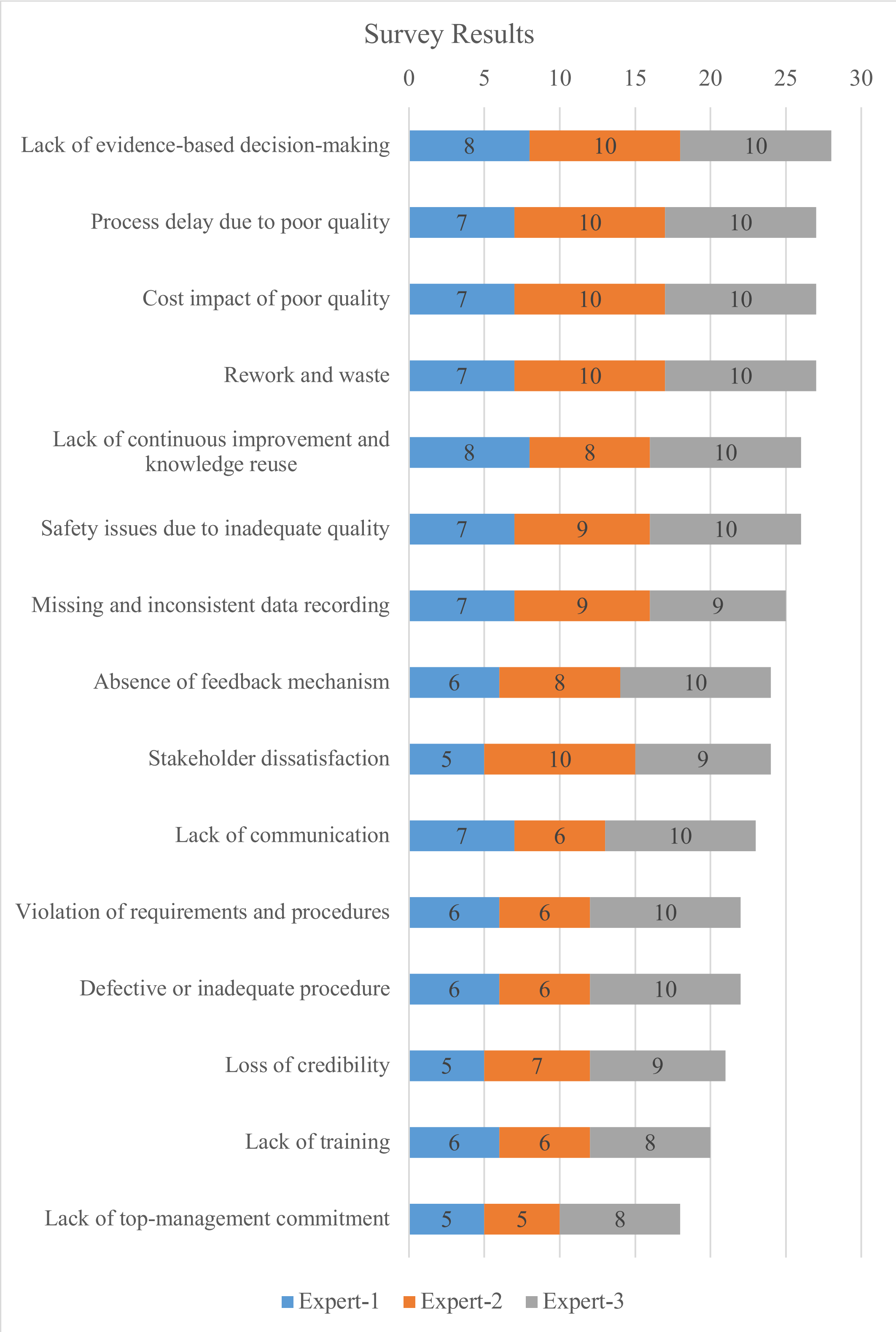

Figure 8.1. Survey Results

CHAPTER 9

CONCLUSION

Quality is a critical aspect of construction; however, projects and processes suffer from lacking quality. Various factors affect the level of quality, and like a domino effect, insufficient quality leads to further problems. Not only do they decrease the overall value of the final product or the process outcome, but also they lead to unintended consequences such as less stakeholder satisfaction, loss of credibility, and so forth. Therefore, there should be a functional and well-designed QMS. However, current approaches are either insufficient or provide only reactive solutions to already-occurred nonconformities, but they cannot prevent them in advance. As a solution, this thesis proposed a QMS that can handle quality issues before and after their occurrence. The statement forming the study's foundation was that there is a strong relationship between KM and QM and that KM can support and nourish QM practices, which was proven by the industrial and academic sources.

Nonetheless, construction companies pay little or no attention to KM applications. Most of them do not develop an organizational memory that contains experience and know-how. This knowledge is only stored in the minds of people, making the organizational memory prone to high turnover. Recording of knowledge is another challenging aspect in this regard. Current processes face insufficient information recording, making the generation of organizational memory impossible. Therefore, most companies are not aware of the value of this tacit knowledge. They have not realized the benefits of failure; instead, they rush to fix what is wrong. Even in the case where nonconformities and lessons learned are recorded and stored, retrieval of this information is not effective and efficient. The methods for reusing this information prevent leveraging its real potential, and time loss while putting effort into this process results in overdue events.

9.1 Contribution and Practical Implications

To enable what the current methods were not able to achieve, an intelligent recommendation tool was contributed by this thesis. By making use of a state-of-the-art language representation model, this tool processes the tacit knowledge generated by quality processes of construction projects. The tacit knowledge in this context was corrections to observed nonconformities, lessons learned from past mistakes, and internal audit findings. By processing this knowledge, it was aimed that future nonconformities could be prevented by making proactive improvements and observed nonconformities could be corrected by recommending learned solutions.

The proposed tool not only addresses the literature gap but also has practical implications for the industry, which was revealed by the survey results. First of all, according to the expert opinion, there is a potential that the tool reduces the probability of making a mistake. When previous lessons are used in current and future projects, the possibility of observing malpractice is expected to decrease. The reason for this is that the tool can substitute an experienced employee making recommendations. Also, the tool can create a continuous improvement cycle in which the improved workflow itself can prevent the chance of such mistakes. Secondly, when the rate of error decreases, consequent cost and schedule impact are also expected to shrink. Therefore, the tool might also have advantages in terms of budget, schedule, and the overall performance of processes. The reduced mistake can also decrease rework and waste, further enhancing project success.

The third implication is building an organizational learning system and organizational memory where the developed KB acts like a brain. In this way, companies can make use of this memory to proactively manage construction quality rather than generating one-time reactive solutions to nonconformities. Also, companies will not have to fully rely on the experience of their personnel. The generated KB can reduce the vulnerability of construction companies against employee turnover. Finally, digitalization and standardization can also be the outcomes of this study, although the study relies on these concepts. The idea and the system that was proposed in this study can encourage others to bring digitalized and standardized practices into their workflow. Ultimately, the study proposes a systemic approach to overall QM strategies, which can lead to a state where zero-mistake is achieved.

9.2 Observed Challenges

The two concepts, digitalization and standardization, should be elaborated further since they are also the challenges for this study. Although the contributions and advantages can have positive impacts on the companies, the designed system relies on digitalization and standardization of construction processes. The knowledge and related data must be stored in a digital environment in order for this tool to work as intended. Having a functional electronic document management system is one of the best enablers for the needed digitalization. In addition to digitalization, standardization is required. The developed tool requires input in a specific format. Therefore, if observations are not recorded in a standard approach, additional effort is required while processing this data. Moreover, without having a standardized flow, the chance of violating the procedures can increase. Finally, both digitalization and standardization may lead to challenges during data recording. The current approach is paper-based, which makes it difficult to process hand-written text. This process may also result in empty data fields that could be important while retrieving the information.

9.3 Limitations and Future Study

In addition to the contributions that the tool might provide and the challenges observed while conducting this research, the limitations of this study are listed below:

Limitation 1: A generic corpus was used to vectorize text in this study. A construction-specific corpus might have been used in the thesis, and the results could have been observed.

Limitation-2: That some of the data were recorded in Turkish and the other portion was recorded in English limited the effective use of the entire KB. A multi-lingual language model could be used to take advantage of the entire dataset. However, this limitation will not constitute any problem if the entire process is standardized and the knowledge is recorded in a single language.

Limitation-3: More attributes could be used to retrieve more relevant information. However, this approach requires a large knowledge set. Once a knowledge set that is rich in quantity and diversity is had, other attributes such as project location, project size, and so forth could improve the retrieval accuracy.

Limitation-4: The KB was not rich enough to retrieve more to-the-point recommendations. More cases in each business line and the department would improve the usability of the tool.

Limitation-5: The evaluation part of the thesis needs to be improved by evaluating how relevant the recommendations are based on further expert opinion.

This study designed a systemic approach to quality issues; however, more effort is required for this system to be functional, and a future study should consider the following improvements. Firstly, the aforementioned limitations should be addressed to provide more relevant solutions to the users of the tool. A construction-specific corpus should be built so that abbreviations and terminologies can be processed more effectively. In addition, other language representation models can be considered and included in the tool, and the performance of each of them should be assessed. This assessment can be done based on expert opinion or other techniques that are found in the literature. To enrich the developed KB, best practices can also be added to retrieve not only the lessons learned from negative events but also those learned from best practices.

The Python code prepared in the thesis will be connected to a user interface (e.g., a web platform) through an application programming interface (API) such as Flask. This will enable performing a case study where the proposed system can be applied to real construction processes. The implementation can be further improved by connecting it to an electronic document management system. This connection between the two software enables the retrieval of information as well as digital documentation. The documentation might include PDFs, spreadsheets, or images. In the future, these different file formats will be processed with certain techniques, such as image recognition. Similarly, input given to the tool can also be diversified in format, and these documents can be automatically categorized with an automated AI-based categorization tool. This feature will increase the practicality of the overall system. Finally, what is explained until now will be connected to a central mechanism that gathers various parts of construction. This central mechanism can enable the monitoring of entire processes from a single node. Setting up this central mechanism and implementing it properly can enable a fully automatized and intelligent construction workflow. Ultimately, the construction industry can be raised to a level where continuous and autonomous improvement and learning culture become the mainstream driver of the industry where zero-mistake culture becomes central.

REFERENCES

Abdul-Rahman, H., Yahya, I. A., Berawi, M. A., & Wah, L. W. (2008). Conceptual delay mitigation model using a project learning approach in practice. *Construction Management and Economics*, *26*(1), 15–27. https://doi.org/10.1080/01446190701744350

Adenfelt, M. (2010). Exploring the performance of transnational projects: Shared knowledge, coordination and communication. *International Journal of Project Management*, *28*(6), 529–538. https://doi.org/10.1016/j.ijproman.2009.10.004

Al Qady, M., & Kandil, A. (2009). Techniques for evaluating automated knowledge acquisition from contract documents. *Construction Research Congress 2009*, 1479–1488. https://doi.org/10.1061/41020(339)150

Al Qady, M., & Kandil, A. (2010). Concept relation extraction from construction documents using natural language processing. *Journal of Construction Engineering and Management*, *136*(3), 294–302. https://doi.org/10.1061/(ASCE)CO.1943-7862.0000131

Ali, A. S., & Rahmat, I. (2010). The performance measurement of construction projects managed by ISO-certified contractors in Malaysia. *Journal of Retail & Leisure Property*, *9*(1), 25–35. https://doi.org/10.1057/rlp.2009.20

Amer, F., Hockenmaier, J., & Golparvar-Fard, M. (2022). Learning and critiquing pairwise activity relationships for schedule quality control via deep learning-based natural language processing. *Automation in Construction*, *134*. https://doi.org/10.1016/j.autcon.2021.104036

Arditi, D., & Murat Gunaydin, H. (1997). Total quality management in the construction process. *International Journal of Project Management*, *15*(4), 235–243. https://doi.org/10.1016/S0263-7863(96)00076-2

Asgari, Z., & Rahimian, F. P. (2017). Advanced virtual reality applications and intelligent agents for construction process optimisation and defect prevention.

Procedia Engineering, *196*, 1130–1137. https://doi.org/10.1016/j.proeng.2017.08.070

Ayhan, B. U., & Tokdemir, O. B. (2019). Safety assessment in megaprojects using artificial intelligence. *Safety Science*, *118*(March), 273–287. https://doi.org/10.1016/j.ssci.2019.05.027

Baek, S., Jung, W., & Han, S. H. (2021). A critical review of text-based research in construction: Data source, analysis method, and implications. In *Automation in Construction* (Vol. 132). Elsevier B.V. https://doi.org/10.1016/j.autcon.2021.103915

Bafna, P., Shirwaikar, S., & Pramod, D. (2019). Task recommender system using semantic clustering to identify the right personnel. *VINE Journal of Information and Knowledge Management Systems*, *49*(2), 181–199. https://doi.org/10.1108/VJIKMS-08-2018-0068

Battikha, M. G. (2002). QUALICON: Computer-based system for construction quality management. *Journal of Construction Engineering and Management*, *128*(2), 164–173. https://doi.org/10.1061/(ASCE)0733-9364(2002)128:2(164)

Bird, S., Klein, E., & Loper, E. (2009). *Nautral Language Processing with Python*. O'Reilly Media Inc.

Bojanowski, P., Grave, E., Joulin, A., & Mikolov, T. (2016). Enriching word vectors with subword information. *ArXiv*. http://arxiv.org/abs/1607.04606

Burati Jr., J. L., Farrington, J. J., & Ledbetter, W. B. (1992). Causes of quality deviations in design and construction. *Journal of Construction Engineering and Management*, *118*(1), 34–49. https://doi.org/10.1061/(ASCE)0733-9364(1992)118:1(34)

Burton, R. M., Obel, Børge., & DeSanctis, Gerardine. (2011). *Organizational Design: A Step-By-Step Approach* (2nd ed.). Cambridge University Press.

Caldas, C. H., Soibelman, L., & Han, J. (2002). Automated classification of construction project documents. *Journal of Computing in Civil Engineering*, *16*(4), 234–243. https://doi.org/10.1061/(ASCE)0887-3801(2002)16:4(234)

Cambraia, F. B., Saurin, T. A., & Formoso, C. T. (2010). Identification, analysis and dissemination of information on near misses: A case study in the construction industry. *Safety Science*, *48*(1), 91–99. https://doi.org/10.1016/j.ssci.2009.06.006

Candaş, A. B., & Tokdemir, O. B. (2022a). Automated identification of vagueness in the FIDIC Silver Book Conditions of Contract. *Journal of Construction Engineering and Management*, *148*(4). https://doi.org/10.1061/(asce)co.1943-7862.0002254

Candaş, A. B., & Tokdemir, O. B. (2022b). Automating coordination efforts for reviewing construction contracts with multilabel text classification. *Journal of Construction Engineering and Management*, *148*(6). https://doi.org/10.1061/(asce)co.1943-7862.0002275

Carrillo, P. (2004). Managing knowledge: Lessons from the oil and gas sector. *Construction Management and Economics*, *22*(6), 631–642. https://doi.org/10.1080/0144619042000226289

Carrillo, P. M., Anumba, C. J., & Kamara, J. M. (2002). *Knowledge management strategy for construction: Key I.T. and contextual issues*. https://itc.scix.net/pdfs/w78-2000-155.content.pdf

Carrillo, P., Robinson, H., Al-Ghassani, A., & Anumba, C. (2004). Knowledge management in UK construction: Strategies, resources and barriers. *Project Management Journal*, *35*(1), 46–56. https://doi.org/10.1177/875697280403500105

Chan, A. P. C., Scott, D., & Chan, A. P. L. (2004). Factors affecting the success of a construction project. *Journal of Construction Engineering and Management*, *130*(1), 153–155. https://doi.org/10.1061/(ASCE)0733-9364(2004)130:1(153)

Chokor, A., Naganathan, H., Chong, W. K., & Asmar, M. el. (2016). Analyzing Arizona OSHA injury reports using unsupervised machine learning. *Procedia Engineering*, *145*, 1588–1593. https://doi.org/10.1016/j.proeng.2016.04.200

Choo, C. W. (2006). *The Knowing Organization: How Organizations Use Information To Construct Meaning, Create Knowledge, and Make Decisions* (2nd ed.). Oxford University Press.

Dharma, E. M., Lumban Gaol, F., Warnars, H. L. H. S., & Soewito, B. (2022). The accuracy comparison among word2vec, glove, and fasttext towards convolutional neural network (CNN) text classification. *Journal of Theoretical and Applied Information Technology*, *31*(2). http://www.jatit.org/volumes/Vol100No2/5Vol100No2.pdf

Ding, Y., Ma, J., & Luo, X. (2022). Applications of natural language processing in construction. *Automation in Construction*, *136*, 104169. https://doi.org/10.1016/j.autcon.2022.104169

Ding, Z., Ng, F., & Cai, Q. (2007). Personal constructs affecting interpersonal trust and willingness to share knowledge between architects in project design teams. *Construction Management and Economics*, *25*(9), 937–950. https://doi.org/10.1080/01446190701468828

Egan, J. (1998). *Rethinking Construction*.

Egbu, C., Botterill, K., & Bates, M. (2001). The influence of knowledge management and intellectual capital on organizational innovations. In A. Akintoye (Ed.), *17th Annual ARCOM Conference* (Vol. 1, pp. 547–555). Association of Researchers in Construction Management. https://www.arcom.ac.uk/-docs/proceedings/ar2001-547-555_Egbu_Botterill_and_Bates.pdf

Egbu, C. O. (1999). Skills, knowledge and competencies for managing construction refurbishment works. *Construction Management and Economics*, *17*(1), 29–43. https://doi.org/10.1080/014461999371808

Egbu, C. O. (2004). Managing knowledge and intellectual capital for improved organizational innovations in the construction industry: An examination of critical success factors. *Engineering, Construction and Architectural Management*, *11*(5), 301–315. https://doi.org/10.1108/09699980410558494

Egwim, C. N., Alaka, H., Toriola-Coker, L. O., Balogun, H., Ajayi, S., & Oseghale, R. (2021). Extraction of underlying factors causing construction projects delay in Nigeria. *Journal of Engineering, Design and Technology*. https://doi.org/10.1108/JEDT-04-2021-0211

Eken, G., Bilgin, G., Dikmen, I., & Birgonul, M. T. (2020). A lessons-learned tool for organizational learning in construction. *Automation in Construction*, *110*. https://doi.org/10.1016/j.autcon.2019.102977

Erfani, A., Cui, Q., & Cavanaugh, I. (2021). An empirical analysis of risk similarity among major transportation projects using natural language processing. *Journal of Construction Engineering and Management*, *147*(12). https://doi.org/10.1061/(ASCE)CO.1943-7862.0002206

Fan, W., Chen, Y., Li, J., Sun, Y., Feng, J., Hassanin, H., & Sareh, P. (2021). Machine learning applied to the design and inspection of reinforced concrete bridges: Resilient methods and emerging applications. *Structures*, *33*, 3954–3963. https://doi.org/10.1016/j.istruc.2021.06.110

FastText. (2020, April 6). Meta AI. https://fasttext.cc/

FastText. (2022, March 13). Facebook Open Source.

Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management*, *35*(2), 137–144. https://doi.org/10.1016/j.ijinfomgt.2014.10.007

Goh, Y. M., & Chua, D. K. H. (2009). Case-based reasoning for construction hazard identification: Case representation and retrieval. *Journal of*

Construction Engineering and Management, *135*(11), 1181–1189. https://doi.org/10.1061/(ASCE)CO.1943-7862.0000093

Goh, Y. M., & Chua, D. K. H. (2010). Case-based reasoning approach to construction safety hazard identification: Adaptation and utilization. *Journal of Construction Engineering and Management*, *136*(2), 170–178. https://doi.org/10.1061/(ASCE)CO.1943-7862.0000116

Grave, E. (2016, August 18). *Releasing fastText*. https://fasttext.cc/blog/2016/08/18/blog-post.html

Grave, E., Bojanowski, P., Gupta, P., Joulin, A., & Mikolov, T. (2018). Learning Word Vectors for 157 Languages. *Proceedings of the International Conference on Language Resources and Evaluation (LREC 2018)*. https://arxiv.org/abs/1802.06893

Grzegorczyk, K. (2018). *Vector representations of text data in deep learning* [Doctoral dissertation, AGH University of Science and Technology]. http://arxiv.org/abs/1901.01695

Hallowell, M. R. (2012). Safety-knowledge management in American construction organizations. *Journal Of Management In Engineering*, *28*(2), 203–211. https://doi.org/10.1061/(ASCE)ME.1943-5479.0000067

Hong, Y., Xie, H., Bhumbra, G., & Brilakis, I. (2021). Comparing natural language processing methods to cluster construction schedules. *Journal of Construction Engineering and Management*, *147*(10). https://doi.org/10.1061/(asce)co.1943-7862.0002165

International Organization for Standardization. (2015). *Quality management systems-Requirements (ISO 9001:2015(E))* (5th ed.). https://www.iso.org/standard/62085.html

Jallan, Y., & Ashuri, B. (2020). Text mining of the securities and exchange commission financial filings of publicly traded construction firms using deep learning to identify and assess risk. *Journal of Construction Engineering and Management*, *146*(12), 04020137. https://doi.org/10.1061/(asce)co.1943-7862.0001932

Jallan, Y., Brogan, E., Ashuri, B., & Clevenger, C. M. (2019). Application of natural language processing and text mining to identify patterns in construction-defect litigation cases. *Journal of Legal Affairs and Dispute Resolution in Engineering and Construction*, *11*(4), 04519024. https://doi.org/10.1061/(asce)la.1943-4170.0000308

Joulin, A., Grave, E., Bojanowski, P., Douze, M., Jégou, H., & Mikolov, T. (2016). FastText.zip: Compressing text classification models. *ArXiv*. https://doi.org/10.48550/arXiv.1612.03651

Joulin, A., Grave, E., Bojanowski, P., & Mikolov, T. (2016). Bag of tricks for efficient text classification. *ArXiv*. https://doi.org/10.48550/arXiv.1607.01759

Kamara, J. M., Anumba, C. J., & Carrillo, P. M. (2002). A CLEVER approach to selecting a knowledge management strategy. *International Journal of Project Management*, *20*(3), 205–211. https://doi.org/10.1016/S0263-7863(01)00070-9

Kim, H., Lee, H.-S., Park, M., Chung, B., & Hwang, S. (2015). Information retrieval framework for hazard identification in construction. *Journal of Computing in Civil Engineering*, *29*(3), 04014052. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000340

Kim, T., & Chi, S. (2019). Accident case retrieval and analyses: Using natural language processing in the construction industry. *Journal of Construction Engineering and Management*, *145*(3), 04019004. https://doi.org/10.1061/(asce)co.1943-7862.0001625

Knowledge Management Research Report. (2000). *KPMG Consulting*. http://www.providersedge.com/docs/km_articles/kpmg_km_research_report_2000.pdf

Kowsher, Sobuj, S. I., Shahriar, F., Prottasha, N. J., Arefin, M. S., Dhar, P. K., & Koshiba, T. (2022). An enhanced neural word embedding model for transfer learning. *Applied Sciences*, *12*(6), 2848. https://doi.org/10.3390/app12062848

Larsen, J. K., Shen, G. Q., Lindhard, S. M., & Brunoe, T. D. (2016). Factors affecting schedule delay, cost overrun, and quality level in public construction projects. *Journal of Management in Engineering*, *32*(1), 04015032. https://doi.org/10.1061/(asce)me.1943-5479.0000391

Lee, J., Yi, J.-S., & Son, J. (2019). Development of automatic-extraction model of poisonous clauses in international construction contracts using rulebBased NLP. *Journal of Computing in Civil Engineering*, *33*(3), 04019003. https://doi.org/10.1061/(asce)cp.1943-5487.0000807

Li, R. Y. M., Li, H. C. Y., Tang, B., & Au, W. C. (2020). Fast AI classification for analyzing construction accidents claims. *ACM International Conference Proceeding Series*, 1–4. https://doi.org/10.1145/3407703.3407705

Liu, C., Sepasgozar, S. M. E., Shirowzhan, S., & Mohammadi, G. (2022). Applications of object detection in modular construction based on a comparative evaluation of deep learning algorithms. *Construction Innovation*, *22*(1), 141–159. https://doi.org/10.1108/CI-02-2020-0017

Love, P. E. D., Edum-Fotwe, F., & Irani, Z. (2003). Management of knowledge in project environments. *International Journal of Project Management*, *21*, 155–156. https://doi.org/10.1016/S0263-7863(02)00089-3

Lu, Y., Li, Q., & Xiao, W. (2013). Case-based reasoning for automated safety risk analysis on subway operation: Case representation and retrieval. *Safety Science*, *57*, 75–81. https://doi.org/10.1016/j.ssci.2013.01.020

Manning, C. D., Surdeanu, M., Bauer, J., Finkel, J., Bethard, S. J., & McClosky, D. (2014). *The Stanford CoreNLP Natural Language Processing Toolkit*. Stanford NLP. https://nlp.stanford.edu/pubs/StanfordCoreNlp2014.pdf

Meziane, F., & Rezgui, Y. (2004). A document management methodology based on similarity contents. *Information Sciences*, *158*(1–4), 15–36. https://doi.org/10.1016/j.ins.2003.08.009

Mikolov, T., Bojanowski, P., Grave, E., & Joulin, A. (2016, August 18). *FAIR open-sources fastText*. Engineering at Meta. https://engineering.fb.com/2016/08/18/ml-applications/fair-open-sources-fasttext/

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient Estimation of Word Representations in Vector Space. *ArXiv*. http://arxiv.org/abs/1301.3781

Mikolov, T., Sutskever, I., Chen, K., Corrado, G., & Dean, J. (2013). *Distributed representations of words and phrases and their compositionality*. *ArXiv*. http://arxiv.org/abs/1310.4546

Mo, Y., Zhao, D., Du, J., Syal, M., Aziz, A., & Li, H. (2020). Automated staff assignment for building maintenance using natural language processing. *Automation in Construction*, *113*(103150). https://doi.org/10.1016/j.autcon.2020.103150

Moon, S., Lee, G., & Chi, S. (2022). Automated system for construction specification review using natural language processing. *Advanced Engineering Informatics*, *51*. https://doi.org/10.1016/j.aei.2021.101495

Nonaka, I., & Takeuchi, H. (1995). *The Knowledge-Creating Company* (1st ed.). Oxford University Press, Inc. ISBN: 0-19-509269-4

Oltra, V. (2005). Knowledge management effectiveness factors: The role of HRM. *Journal of Knowledge Management*, *9*(4), 70–86. https://doi.org/10.1108/13673270510610341

Pathirage, C. P., Amaratunga, D. G., & Haigh, R. P. (2007). Tacit knowledge and organisational performance: Construction industry perspective. *Journal of*

Knowledge Management, *11*(1), 115–126. https://doi.org/10.1108/13673270710728277

Pheng, L. S., & Chuan, Q. T. (2006). Environmental factors and work performance of project managers in the construction industry. *International Journal of Project Management*, *24*(1), 24–37. https://doi.org/10.1016/j.ijproman.2005.06.001

Poh, C. Q. X., Ubeynarayana, C. U., & Goh, Y. M. (2018). Safety leading indicators for construction sites: A machine learning approach. *Automation in Construction*, *93*, 375–386. https://doi.org/10.1016/j.autcon.2018.03.022

Ren, R., & Zhang, J. (2021). Semantic rule-based construction procedural information extraction to guide jobsite sensing and monitoring. *Journal of Computing in Civil Engineering*, *35*(6), 04021026. https://doi.org/10.1061/(asce)cp.1943-5487.0000971

Robinson, H. S., Carrillo, P. M., Anumba, C. J., & Al-Ghassani, A. M. (2004). Developing a business case for knowledge management: The IMPaKT approach. *Construction Management and Economics*, *22*(7), 733–743. https://doi.org/10.1080/0144619042000226306

Robinson, H. S., Carrillo, P. M., Anumba, C. J., & Al-Ghassani, A. M. (2005). Knowledge management practices in large construction organisations. *Engineering, Construction and Architectural Management*, *12*(5), 431–445. https://doi.org/10.1108/09699980510627135

Salama, D. M., & El-Gohary, N. M. (2011). Semantic modeling for automated compliance checking. *International Workshop on Computing in Civil Engineering 2011*, 641–648. https://doi.org/10.1061/41182(416)79

Shin, S., & Issa, R. R. A. (2021). BIMASR: Framework for voice-based BIM information retrieval. *Journal of Construction Engineering and Management*, *147*(10), 04021124. https://doi.org/10.1061/(asce)co.1943-7862.0002138

Siemieniuch, C. E., & Sinclair, M. A. (1999). Organizational aspects of knowledge lifecycle management in manufacturing. *International Journal of Human-Computer Studies*, *51*(jhc.1999.0274), 517–547. https://doi.org/10.1006/ijhc.1999.0274

Sik-wah Fong, P., & Chu, L. (2006). Exploratory study of knowledge sharing in contracting companies: A sociotechnical perspective. *Journal of Construction Engineering and Management*, *132*(9), 928–939. https://doi.org/10.1061/(ASCE)0733-9364(2006)132:9(928)

Singhal, A. (2001). *Modern information retrieval: A brief overview*. http://singhal.info/ieee2001.pdf

Soibelman, L., & Kim, H. (2002). Data preparation process for construction knowledge generation through knowledge discovery in databases. *Journal of Computing in Civil Engineering*, *16*(1), 39–48. https://doi.org/10.1061/(ASCE)0887-3801(2002)16:1(39)

Styhre, A., & Gluch, P. (2010). Managing knowledge in platforms: Boundary objects and stocks and flows of knowledge. *Construction Management and Economics*, *28*(6), 589–599. https://doi.org/10.1080/01446190903450061

Tang, L., Griffith, L., Stevens, M., & Hardie, M. (2020). Social media analytics in the construction industry comparison study between China and the United States. *Engineering, Construction and Architectural Management*, *27*(8), 1877–1889. https://doi.org/10.1108/ECAM-12-2019-0717

Tixier, A. J. P., Hallowell, M. R., Rajagopalan, B., & Bowman, D. (2016). Automated content analysis for construction safety: A natural language processing system to extract precursors and outcomes from unstructured injury reports. *Automation in Construction*, *62*, 45–56. https://doi.org/10.1016/j.autcon.2015.11.001

Tserng, H. P., & Lin, Y. C. (2004). Developing an activity-based knowledge management system for contractors. *Automation in Construction*, *13*(6), 781–802. https://doi.org/10.1016/j.autcon.2004.05.003

Tupenaite, L., Kanapeckiene, L., Naimaviciene, J., Tupenaite, L., Kanapeckiene, L., & Naimaviciene, J. (2008). Knowledge management model for construction projects. *Computer Modelling and New Technologies*, *12*(3), 38–46. https://www.researchgate.net/publication/228818669

Wang, N., Issa, R. R. A., & Anumba, C. J. (2022). NLP-based query-answering system for information extraction from building information models. *Journal of Computing in Civil Engineering*, *36*(3). https://doi.org/10.1061/(asce)cp.1943-5487.0001019

Wirth, R., & Hipp, J. (2000). CRISP-DM: Towards a standard process model for data mining. *Proceedings of the 4th International Conference on the Practical Applications of Knowledge Discovery and Data Mining*, 29–40. http://www.cs.unibo.it/~danilo.montesi/CBD/Beatriz/10.1.1.198.5133.pdf

Yang, L. R., Chen, J. H., & Wang, H. W. (2012). Assessing impacts of information technology on project success through knowledge management practice. *Automation in Construction*, *22*, 182–191. https://doi.org/10.1016/j.autcon.2011.06.016

Yepes, V., & López, S. (2021). Knowledge management in the construction industry: Current state of knowledge and future research. *Journal of Civil Engineering and Management*, *27*(8), 671–680. https://doi.org/10.3846/jcem.2021.16006

Zhang, J., & El-Gohary, N. (2012a). Automated regulatory information extraction from building codes leveraging syntactic and semantic information. *Construction Research Congress 2012*, 622–632. https://doi.org/10.1061/9780784412329.063

Zhang, J., & El-Gohary, N. (2012b). Extraction of construction regulatory requirements from textual documents using natural language processing techniques. *International Conference on Computing in Civil Engineering*, 453–460. https://doi.org/10.1061/9780784412343.0057

Zhu, H., Hwang, B.-G., Ngo, J., & Tan, J. P. S. (2022). Applications of smart technologies in construction project management. *Journal of Construction Engineering and Management*, *148*(4). https://doi.org/10.1061/(ASCE)CO.1943-7862.0002260

Zou, Y., Kiviniemi, A., & Jones, S. W. (2017). Retrieving similar cases for construction project risk management using natural language processing techniques. *Automation in Construction*, *80*, 66–76. https://doi.org/10.1016/j.autcon.2017.04.003

Zulkifli, A. R., Yahya, N. A., Azmi, E. A., & Diah, J. M. (2018). Site inspection practice in construction projects in Kuala Lumpur, Malaysia. *AIP Conference Proceedings*, *2020*, 020050. https://doi.org/10.1063/1.5062676

APPENDICES

A. Detailed Integration into Overall Quality Management Strategy

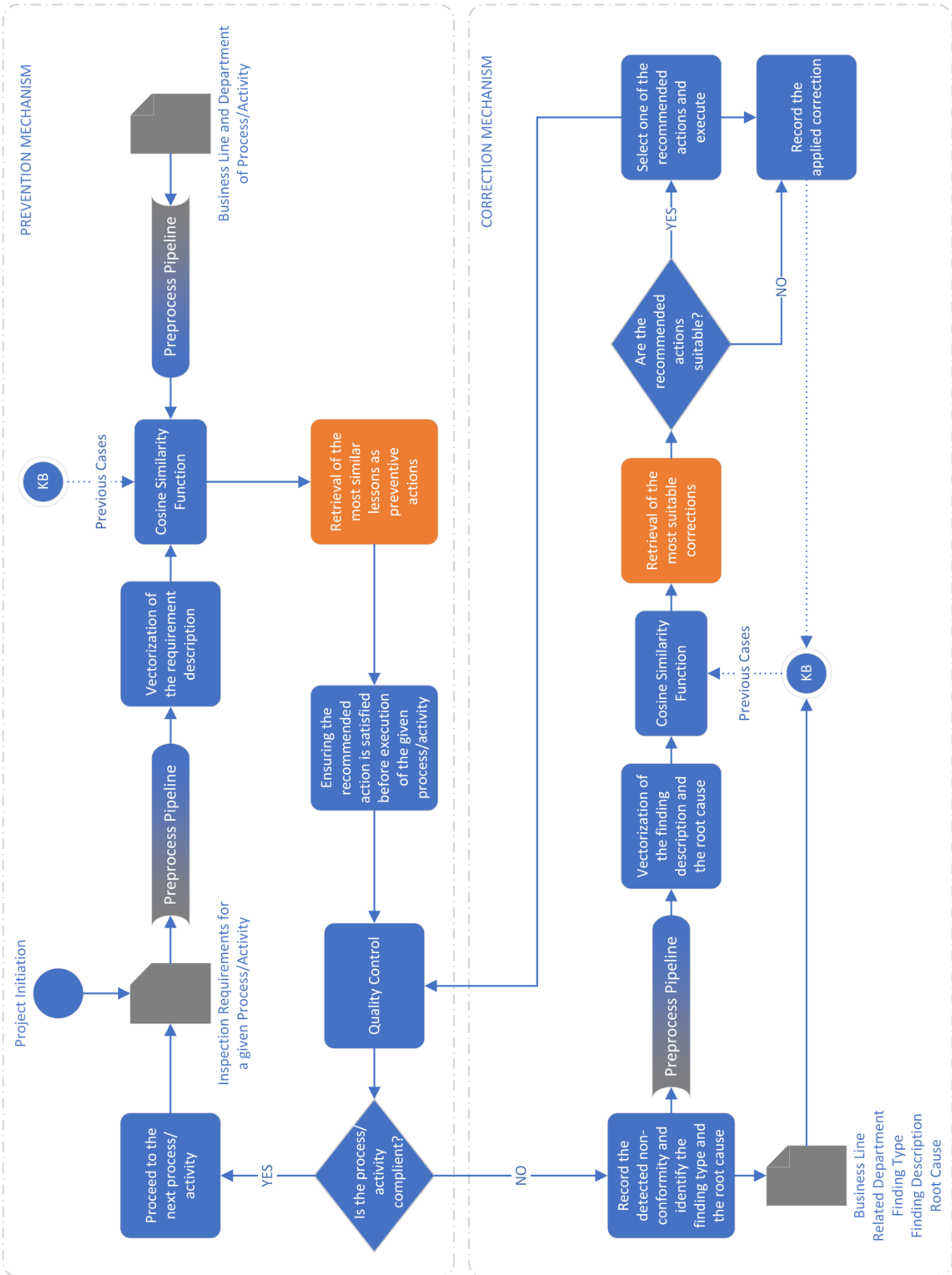

Figure A.1. Detailed Integration Into Overall Quality Management Strategy