A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy in Architecture

A Text Mining Approach to Performance Enhancement in BIM Pervasive Major Project Delivery

Ambark Bareka
PhD Researcher
Registration No 201887253
Department of Architecture

2024

Declaration

This thesis is the result of the author's original research. It has been composed by the author and has not been previously submitted for examination which has led to the award of a degree.

Signed:

Date:

COVID-19 Impact Statement

In accordance with the university's policy "Covid Impact Statement for PGR Thesis Submissions" issued by the Strathclyde Doctoral School and approved by the University Senate in 2021, this statement is made to explain the impacts to and adjustments made to my research in response to the significant disruptions caused by the COVID-19 pandemic.

This impact statement includes the following three parts required by this university polity:

Requirement 1) *A brief overview of the pre-Covid-19 research plan including details such as research design, methods, analytical methods where appropriate to your discipline and how these connect within the thesis plan.*

Statement:

Before the onset of the Covid-19 pandemic, my research was designed to investigate the integration of Text Mining (TM) with Building Information Modelling (BIM) across multiple case studies. The initial research plan included fieldwork at various project sites, employing qualitative methods such as semi-structured interviews with key stakeholders, direct observations, and the collection of primary data from surveys and project documentation. These methods were intended to provide rich, contextual insights into the practical application of TM within BIM environments.

The analytical framework combined thematic coding and qualitative content analysis with computational text mining techniques, facilitating the extraction and comparison of patterns across diverse datasets. This multi-case study approach was central to the thesis structure, with chapters dedicated to the literature review, methodology, findings from individual case studies, cross-case analysis, and conclusions. The aim was to develop a comprehensive understanding of the potential for TM to enhance decision-making processes within BIM projects.

Requirement 2) *Details on how Covid-19 related disruption impacted your research (for example, difficulties collecting/analysing data because of travel/lab access restrictions, caring responsibilities, or health issues).*

Statement:

The Covid-19 pandemic presented significant challenges to my research, particularly due to restrictions on travel, site access, and face-to-face engagement. Fieldwork, a cornerstone of my research design, became unfeasible as lockdowns and social distancing measures were introduced. Planned site visits were cancelled, and in-person interviews with project stakeholders were either postponed indefinitely or moved to virtual platforms, which limited the depth and quality of interaction.

In addition to these logistical disruptions, my personal circumstances during the pandemic had a notable impact. As the primary caregiver for my wife and young son, I had to balance household responsibilities and provide both practical and emotional support while continuing my academic work. This dual burden constrained the time and focus I could dedicate to research, compounding the delays caused by the pandemic.

Institutional closures further hindered my ability to access physical resources, such as archives and technical documentation. Adapting to these constraints while working remotely required significant adjustments to both the timeline and methodology of my research.

Requirement 3) *A summary of any decision taken to mitigate for Covid-19 restrictions in terms of reduced datasets or shorter chapters, changes to topic, methods, location of research etc.*

Statement:

To mitigate the impacts of these disruptions, I made substantial adjustments to my research design. The scope of the study was narrowed to focus on a single case study: the Crossrail project. This project was selected due to the availability of comprehensive secondary data, including technical documentation and project reports, which could be accessed remotely. This adjustment allowed me to progress my research without requiring site visits or in-person interactions.

The shift in scope enabled a deeper exploration of the integration of TM with BIM within the specific context of the Crossrail project. I adapted computational text mining techniques to analyse the rich textual data available, maintaining methodological rigour and ensuring that the research questions could still be addressed meaningfully. While the original multi-case study approach was not feasible, the single case study provided sufficient depth and focus to deliver valuable insights into the research topic.

These adjustments were made in accordance with the University of Strathclyde's policy on Covid-19 impact statement, which emphasised the need for flexibility in research approaches during the pandemic. By adapting my research plan in this way, I was able to ensure the feasibility and relevance of the study despite the unprecedented challenges posed by Covid-19.

This statement highlights the changes made to the research methodology and design in response to the challenges faced during the pandemic, as advised by the University policy "Covid Impact Statement for PGR Thesis Submissions" (2021).

PhD Candidate: Ambark Bareka

Lead Supervisor: Dr Zhen Chen

Acknowledgements

First of all, I wish to express my heartfelt gratitude to the Libyan Embassy for their gracious sponsorship of this esteemed PhD scholarship. Their generous support not only aids in the pursuit of academic excellence but also symbolizes a significant step towards my personal and professional growth.

Numerous individuals have played integral roles in the development of this work. Foremost, I express my sincere gratitude to Dr Zhen Chen for accepting the responsibility of serving as my supervisor. His invaluable insights, constructive criticisms, and unwavering support have been pivotal in guiding this research to its completion. Additionally, I wish to acknowledge the significant contributions of Dr Andrew Agapiou, who has served as a dedicated second supervisor. His commitment and efforts have greatly facilitated the execution of this study, for which I am deeply appreciative.

During my years of study, Zhen Chen was my supervisor. He had a deep understanding of syntax and was really sharp. Some of the ideas in this work are based on his guidance and support, and I'm really thankful for his contribution. I also want to thank Professor Branka Dimitrijevic. She took part in several conferences that added a lot to this research. I appreciate Professor Ashraf Salama (who is now at Northumbria University) and all the staff in the Architecture Department. They were kind and supportive during my time of study. I'm also grateful to the librarians who were always friendly and ready to help whenever I needed it.

I would also like to extend a special thanks to my examiners, Professor Bimal Kumar at the University of Strathclyde and Professor Heng Li at the Hong Kong Polytechnic University, for their time, effort, and invaluable feedback. Their expertise and insights have been essential to refining and finalising this research.

Lastly, big thanks to my mother, brothers, and relatives back at home. They kept supporting and encouraging me, and I'm truly thankful for that. Special appreciation also goes to my wife and kids who were with me throughout this journey. Their unwavering support has been invaluable.

Abstract

Purpose: This research aims to rigorously develop and implement a specialized TM technique to significantly enhance performance in BIM pervasive major project delivery. By addressing challenges related to errors, uncertainties, and expectations within this context, the study endeavours to offer substantial contributions to construction management knowledge. The research introduces the "BIM Text Analysis and Performance Enhancement" (BIMTAPE) framework, which synergizes the TM methodology with BIM to optimize project execution within the BIM pervasive major project delivery.

Methodology: A holistic systematic literature review enabled by the Nine-Square Process (NSP) was conducted, entailing a thorough search of relevant literature from academic databases and other relevant sources, identifying current knowledge and the research gap on delivery. Additionally, the research process adhered to a sequenced approach, where the development of the BIMTAPE Framework took precedence before conducting specific experimental case studies. This structured framework provided a systematic and comprehensive approach to analyse diverse BIM-related texts.

Findings: This research validates experimental findings and provides insights into the relationship between text analysis and performance enhancement in major project delivery, using Social Network Analysis. The study comprehensively understands the role of text analysis in major project performance, emphasizing BIM alignment with industry standards for value realization in contracts. Notably, the BIMTAPE Framework's potential impact in addressing challenges, exemplified by the High Speed Two (HS2) project's initiatives, enhances major project execution within the BIM pervasive major project delivery.

Implications: The integration of BIM technology with TM within the BIMTAPE Framework excludes structured data standards like Industry Foundation Classes (IFC) and buildingSMART Data Dictionary (bSDD). In academic settings, the framework introduces a new approach for investigating unstructured textual data in BIM, encouraging further exploration.

Limitations: This research's applicability is constrained by the limited sample size. Expanding the scope to encompass larger studies will be crucial for generalisability. Further exploration across various contexts will provide a more comprehensive understanding of the framework's implications.

Contents

Figures

Tables

Abbreviations

3D	Three Dimensional
5D	Five Dimensional
AEC	Architecture, Engineering, and Construction
AI	Artificial Intelligence
ANN	Artificial Neural Network
ARL	Association Rule Learning
ASUM-DM	Analytics Solutions Unified Method for Data Mining
BEPP	Building Energy Performance Prediction
BIM	Building Information Modelling
BIMTAPE	BIM Text Analysis and Performance Enhancement Framework
BoW	Bag of Words
BSI	British Standard Institution
CAD	Computer Aided Design
CMBOK	Construction Management Body of Knowledge
CRISP-DM	Cross Industry Standard Process for Data Mining
CSS	Cascading Style Sheets
DBB	Design Bid Build
DCMS	Digital Culture Media and Sport
DIK	Data, Information, and Knowledge
DMAT	Digital Modelling and Associated Technologies
DTM	Document-Term Matrix
EBL	Evidence-Based Learning
EMG	Environmental Management Guideline
FHWA	Federal Highway Administration
FL	Fuzzy Logic
GCS	Government Construction Strategy
GIS	Geographic Information System
HTML	Hypertext Markup Language
ICE	Institute of Civil Engineers
IFC	Industry Foundation Classes
IoT	Internet of Things
ISO	International Organization for Standardization

KDD	Knowledge Discovery in Databases
KPIs	Key Performance Indicators
LDA	Latent Dirichlet Allocation
ML	Machine Learning
MoMs	Minutes of Meetings
NBS	National Building Specification
NLP	Natural Language Processing
NSP	Nine-Square Process
OCR	Optical Character Recognition
PAS	Publicly Available Specification
PDF	Portable Document Format
RFI	Request For Information
RIBA	Royal Institute of British Architects
ROI	Return of Investment
SCA	Solution Content and Applications
SCM	Supply Chain Management
SEMMA	Sample, Explore, Modify, Model, and Assess
SNA	Social Network Analysis
SVM	Support Vector Machines
TAPE	Text Analysis and Process Enhancement
TC	Total Citation
TDM	Term Document Matrix
TF-IDF	Term Frequency-Inverse Document Frequency
TRIZ	Theory of Inventive Problem Solving
TS	Text Set
URL	Uniform Resource Locator
XHTML	EXtensible Hypertext Markup Language

Glossary

Artificial Intelligence (AI): The technology that allows machines and computer systems to replicate human cognitive abilities such as learning, reasoning, problem-solving, and decision-making. AI systems can analyse data, recognise patterns, understand and respond to natural language, and make informed decisions based on new information and experiences. These systems can operate autonomously, performing tasks without the need for human intervention. In construction and project management, AI technologies, such as machine learning and deep learning, are used to optimise workflows, automate decision-making, and improve overall project efficiency (IBM, 2024). AI is discussed in Section 2.8 to explain its applications in project management, decision-making, and optimization within the BIM context.

BIM Knowledge Graph (BIMKG): A specialised type of Knowledge Graph (KG) tailored for BIM. In BIM, it systematically organises and represents information on building elements, project details, relationships, and relevant data. It acts as a structured and interconnected knowledge repository, enhancing understanding and utilisation in the BIM domain. The BIMKG enables effective navigation, query, and analysis of BIM data, improving decision-making, collaboration, and project management in construction and architecture (Liu et al., 2016). BIMKG is illustrated in Section 2.10.1.2 to show how it complements the BIMTAPE Framework by enhancing knowledge accessibility and integration in BIM project delivery.

BIM pervasive project: A construction or built environment project where BIM is extensively adopted among project stakeholders in all project-related professional areas.

BIM pervasive project delivery: The delivery of construction or built environment projects in BIM pervasive work process environment (ISO, 2021).

Case Based Reasoning (CBR): A problem-solving paradigm that involves solving new problems based on similar past cases, where a case library is used to store and retrieve past cases, and adaptation and evaluation processes are used to solve new problems (Kolodner, 2014).

Decision Support Systems (DSS): A computer-based system that provides information and support for decision-making in complex, unstructured or semi-structured problems, where data analysis, modelling, and visualization tools are used to support decision-making (Power, 2018).

Expert Systems (ES): A computer-based system that emulates the decision-making ability of a human expert in a specific domain, where knowledge is represented in a knowledge base and inference rules are used to make decisions (Russell and Norvig, 2010).

Knowledge Graph (KG): The structured representations of knowledge that incorporate entities, relationships, and attributes to depict information in a graph-like format. They organise data in a way that reflects the relationships and connections between different entities, allowing for a more contextual and interconnected understanding of information. KGs are often used to represent and query complex knowledge in a machine-readable format, facilitating the extraction of meaningful insights and the discovery of relationships within a dataset (Fensel et al., 2020). KGs are explained in Section 2.10.1 to describe how they improve data representation, contextual reasoning, and knowledge discovery in BIM projects.

Major Construction Project (MCP): The Chartered Institute of Building (CIOB) introduced the CIOB Planning Protocol 2021 (CIOB PP21), a technical information sheet published in February 2021, which outlines a comprehensive characterization of a major project. According to the CIOB PP21, a major project pertains to endeavours wherein the projected expenses for construction activities are anticipated to surpass £10 million pounds sterling (CIOB Planning protocol, 2021).

Project Performance: The comprehensive assessment of whether a project has achieved its scope, cost, and schedule objectives. Regular evaluations conducted during project monitoring and control phases to observe project execution and detect deviations from the Project Management Plan, enabling proactive measures for mitigation (McFadden et al., 2003).

Support Vector Machines (SVM): A machine learning algorithm used for classification and regression analysis that finds the optimal hyperplane that separates the data points into different classes or predicts the continuous value of the target variable (Bishop, 2006; Schölkopf and Smola, 2002).

The Periodic Table of BIM: The Periodic Table of BIM by NBS is a visual guide inspired by the traditional periodic table. It serves as a reference tool for those involved in BIM implementation, outlining key terms and concepts essential for collaboration in BIM processes. This table documents stages for effective collaboration, encompassing technology, standards, and enabling tools. Unlike the original periodic table, it is flexible, organized into nine groupings. It acts as a handy reference for BIM readiness, suitable for printing or digital sharing (Mordue, 2016).

TM informed BIM: It refers to using TM techniques to enhance BIM by extracting insights from unstructured textual data. It enables data-driven decision-making and improves project management by revealing patterns and relationships within textual content.

Publications from This PhD Research

This section presents a compilation of scholarly contributions arising from the research conducted as part of the PhD journey. The publications encompass a diverse range of topics, each contributing to the advancement of knowledge within the realm of BIM pervasive major project delivery and architectural education. The following list showcases the published works:

Manuscript

Bareka, A., Chen, Z., Agapiou, A., and Dimitrijevic, B. (2021). A text analysis roadmap for BIM pervasive major project delivery. *Frontiers in Built Environment*. Manuscript under internal review.

Journal Papers

Zaed, O., Chen, Z., and Bareka, A. (2022). A future-proof curriculum framework for BIM excellence in architectural education. In *Innovations*, *Disruptions and Future Trends in the Global Construction Industry*. 25/11/2022. Routledge, UK.

Conference Papers

1. Bareka, A., Chen, Z., Agapiou, A., and Zaed, O. (2022). A text mining approach to performance enhancement in BIM pervasive major project delivery. *Proceedings of the International Postgraduate Research Conference (IPGRC)*. 04-06/04/2022, University of Salford, Salford, UK.

2. Bareka, A., Chen, Z., Agapiou, A., and Dimitrijevic, B. (2021). A lifecycle social network analysis framework for BIM adoption in major construction projects. *Proceedings of the 5th Coventry University Construction Conference*. 17/03/2021, Coventry, UK.

3. Bareka, A., Chen, Z., Agapiou, A., and Dimitrijevic, B. (2021). A text mining approach to trend analysis on BIM research in the Middle East and North Africa (MENA). Proceedings of the 5th Coventry University Construction Conference. 17/03/2021, Coventry, UK.

4. Bareka, A., and Chen, Z. (2019). A lifecycle-oriented SNA framework for BIM adoption in major construction project. *Proceedings of the Doctoral Workshop on Contemporary Advances in Research Methodology in Construction Management (ARCOM*). 01/11/2019, Glasgow, UK.

5. Zaed, O., Chen, Z., Grant, M., Dimitrijevic, B., and Bareka, A. (2021). Developing a BIM integrated curriculum framework for undergraduate architectural education in Libya. *Proceedings of the 2nd National Conference for Developing Higher Education Institutions in Libya*. 07/10/2021, Bani Walid University, Bani Walid, Libya.

6. Zaed, O., Chen, Z., and Bareka, A. (2022). An educational contribution to applicable digital twins through BIM integrated curriculum development in architecture. *Proceedings of the International Postgraduate Research Conference (IPGRC)*. 04-06/04/2022, University of Salford, Salford, UK.

Datasets

1. Bareka, A. (2022) *A Comprehensive Academic Articles Dataset Assembly for Analysis*. University of Strathclyde, V1, https://doi.org/10.17632/yk7vgxtjgp.1

2. Bareka, A. (2023), *Crossrail Project Dataset for In-Depth Analysis*. University of Strathclyde, V3, https://doi: 10.17632/5hcg7tddc4.3

Chapter 1
Introduction

1.1 Research Background and Motivations

Building construction methods and technology have evolved rapidly through time through the incorporation of technologies like Artificial Intelligence (AI). One of such technologies is Building Information Modelling (BIM), which is a model-based process that helps in designing, planning, managing, and constructing of buildings (Naamane and Boukara, 2015). BIM has rapidly gained popularity in sustainable building design and enables the engagement of all project stakeholders in the lifecycle analysis of building design. However, BIM is still in its nascent stages, and must be evaluated for its qualitative and quantitative contribution in the performance enhancement of building construction projects, especially for those projects that experience difficulties on key performance issues (Habibi et al., 2019) including budget, quality, and schedule.

BIM produces a digital representation of a building or infrastructure project, which can be used to improve the performance of construction projects by enhancing aspects such as efficiency, accuracy, cost-effectiveness, and sustainability throughout the project lifecycle (Azhar et al., 2009), particularly for BIM pervasive project delivery.

The term "BIM pervasive project delivery" used in this study refers to an approach where BIM is integrally embedded in all phases of a construction or built environment project, creating a collaborative, BIM-driven work process environment (ISO, 2021) that allows continuous data exchange and supports a comprehensive application of BIM’s capabilities throughout the lifecycle of construction or built environment projects. Importantly, "BIM pervasive project" refers to a construction or built environment project where BIM is extensively utilised across all professional areas involved, ensuring that BIM tools and methods are universally adopted by project stakeholders. By distinguishing scopes and characteristics between the two new terms and their interconnections, this study underscores the potential for BIM to drive performance improvements in the delivery of complex, large-scale construction or built environment projects, i.e. major projects, grounded in a pervasive and integrated BIM approach.

Text Mining (TM) is a technique used to extract information from textual data, such as reports or documents, which can be used in conjunction with BIM to enhance its effectiveness. For example, TM can be used to extract information from project documents and add it to the BIM model, making it more accurate and informative. TM has been increasingly used in recent years as a technique to extract and analyse information from project documents in the construction industry, including for BIM projects. Research has shown that TM can be used to extract information from project documents such as contracts, design drawings, and specifications, and add it to the BIM model to improve its accuracy and completeness. This can help to identify potential issues or conflicts between different parts of the project and aid in decision making (Teng et al., 2022).

TM can also be used to analyse project communications, such as emails and meeting minutes, to identify patterns and trends in project performance, detect potential risks, and improve team collaboration. There are also studies that have implemented TM to extract information related to building components, systems, and equipment, which can be useful to optimize building performance, reduce energy consumption, and improve maintenance (Teng et al., 2022). Additionally, TM can be used to extract data from large amounts of unstructured data (Wankhade et al., 2022), such as social media, blogs, news articles, etc., to identify relevant topics, sentiment, and other information that may be useful for decision-making in construction projects.

In general, recent studies, according to Guo et al. (2022) the use of TM in BIM projects is still an active area of research, and there is a growing interest in developing methods and tools that can be applied to enhance the performance of construction projects. The motivations for research in this area includes the need to improve the performance of construction projects, the increasing amount of unstructured data available in construction projects, and the growing use of BIM in the construction industry. Construction projects are inherently complex, and involve many stakeholders, making it difficult to manage and control them for maximum efficiency. TM can be used to extract information from project documents and communications to improve project performance by providing insights into project status, identifying potential risks, and improving communication among stakeholders.

Another motivation for research in this area is the increasing amount of unstructured data available in construction projects. Project documents and communications, such as emails and meeting minutes, are often unstructured and difficult to analyse. TM techniques can be used to extract information from this data and make it more useful for decision-making. Finally, the growing use of BIM in the construction industry is also a motivation for research in this area. BIM is becoming an increasingly important tool for managing construction projects, and TM can be used to improve the accuracy and completeness of BIM models by extracting information from project documents and communications.

Overall, the background and motivations for research on TM approach for performance enhancement in BIM pervasive major project delivery are to improve the performance of construction projects through the use of BIM by extracting and analysing information from project documents and communications using TM techniques.

1.1.1 Knowledge Gap

In the context of researching how TM can enhance performance in BIM-based construction projects, a gap analysis could focus on identifying existing limitations and challenges of applying TM in construction, along with pinpointing areas where further research is essential to leverage TM and BIM for improving major construction project performance.

1.1.2 Research Scope

This research focuses on identifying and evaluating TM techniques that can be used to extract information from project documents, communicate and integrate this information into the BIM model, and improve the performance of construction projects. The study reviews existing literature on TM in construction projects, analysing case studies of TM in construction projects, and evaluating TM tools and techniques currently available for use in such projects. The research scope also encompasses the development of a methodology for integrating TM into the BIM process, as well as the testing and validation of this methodology through experimental case studies. It focuses on specific areas of BIM-based construction projects wherein TM can have the greatest impact on performance, such as building performance, energy consumption, and maintenance.

The gap analysis and research scope of this study aims to identify the current limitations and challenges of using TM in construction projects, as well as the areas where further research is needed, and to develop and test a methodology for integrating TM into the BIM process to improve the performance of BIM-based construction projects.

1.1.3 Research Questions

Although the value of BIM adoption has been extensively studied (Chen et al., 2019) in relation to its strong facilitation potential for construction project management (Hardin and McCool, 2015; Liu et al., 2017; Succar, 2009), it is still under qualitative and quantitative evaluation in terms of performance enhancement through quantifying its effectiveness and efficiency in major construction projects (Chan and Ejohwomu, 2018). Consequently, such projects can commonly experience performance issues as assayed by Key Performance Indicators (KPIs), encompassing factors such as budget, quality, and schedule, etc. The following research questions are therefore identified to develop a TM approach to BIM integrated performance enhancement in major construction project delivery:

- Question 1: Which TM techniques can be used to extract information from project documents?
- Question 2: How effective are the TM techniques in enhancing the performance of major construction projects?
- Question 3: How can TM techniques help to address the limitations of current BIM adoption in major construction projects?
- Question 4: What are the potential benefits in terms of better decision-making, cost savings, increased efficiency, and improved project outcomes?
- Question 5: What are the key challenges and limitations associated with the integration of TM approaches into BIM models, and how can these be

addressed to ensure successful implementation in major construction projects?

In the attempt to find answers to these research questions, this research therefore focuses on identifying and evaluating TM techniques that can be used to extract important information from project documents and communications so as to integrate such valuable information into the BIM pervasive processes to improve project delivery performance in major construction or built environment projects, i.e. major capital projects or major projects as called in this research.

1.2 Research Aim and Objectives

Based on the research background and motivations, the aim of this PhD research is to establish a TM approach to performance enhancement in BIM pervasive major project delivery.

Research objectives

In order to achieve this overall research aim, the study has set up three specific research objectives listed below. These objectives act as important research steps or milestones, helping the research move forward and achieve its aim at creating a novel TM approach to improving the performance of BIM pervasive major project delivery. The three research objectives include

- Research Objective 1: To conduct a systematic literature review to justify the need for integrating BIM and TM for project performance enhancement.
- Research Objective 2: To develop a TM methodology and an accompanied BIMTAPE framework for enhanced performance in BIM pervasive major project delivery.
- Research Objective 3: To justify the effectiveness of the new methodology for major project delivery in terms of performance enhancement.

The first objective is to establish a rationale for the study by presenting the results of an extensive literature review on the topic, which sought to analyse all previous scholarly works and articles pertinent to the subject, to provide a strong justification for the research aim of developing a TM approach for BIM-integrated performance enhancement in major construction project delivery. Reviewing the available literature provides a comprehensive understanding of the current state of the art and the gaps in the existing knowledge. It also helps to identify the most promising TM techniques and tools that can be integrated into the BIM process, as well as the specific areas of BIM-based construction projects where TM can have the greatest impact on performance such as cost, time, and quality.

The second objective is to develop a TM methodology and an accompanied BIMTAPE Framework for BIM pervasive major project delivery, entailing the following steps:

- Step 1: Identify the key information that needs to be extracted from project documents and communications and integrated into the BIM model. This could include information on project scope, schedule, budget, quality, and safety.
- Step 2: Select the most appropriate TM techniques and tools for extracting and analysing the identified information.
- Step 3: Develop a comprehensive data pre-processing and cleaning strategy to ensure that the extracted information is accurate and complete.
- Step 4: Integrate the TM techniques and tools into the BIM process, to improve the accuracy and completeness of the BIM model.
- Step 5: Develop the BIMTAPE Framework that guides the integration of TM techniques into the BIM process and defines how the extracted information will be used to improve the performance of the construction project.
- Step 6: Test the TM methodology and BIMTAPE Framework using Social Network Analysis (SNA) to evaluate its effectiveness and efficiency in enhancing the performance of major construction projects.
- Step 7: Refine the TM methodology and the BIMTAPE Framework based on the results of the SNA.
- Step 8: Disseminate the results and proposed methodology and BIMTAPE Framework through publications and presentations to the academic and professional community.

It is worth noting that the development of a TM methodology and BIMTAPE Framework is an iterative process, with constant refinement and testing based on the SNA results. By applying SNA to the data extracted from project documents, the research can evaluate how the TM methodology and BIMTAPE Framework are impacting the performance of the construction project. The methodology and framework should be flexible enough to adapt to the unique characteristics of different construction projects and should be able to handle the complexity and diversity of the data in construction projects.

The third research objective is to justify the use of TM as a relatively novel methodology, yet to be tested for its efficiency. This can be achieved by examining the literature on the topic of TM to throw light on its purported merits, which include the following:

- **Time-saving:** TM allows for the automatic and efficient extraction of relevant information from large volumes of unstructured data, such as project documents and communications. This saves time and effort compared to manual searching, which can be time-consuming and prone to human error.
- **Relevance:** TM uses Natural Language Processing (NLP) techniques to find patterns and pull out texts that are directly relevant to the topic. However, this process can be complex, as NLP needs to handle different language styles, ambiguities, and context shifts. Advanced algorithms help reduce irrelevant or repetitive information, making the extraction process more accurate and efficient.
- **Multilingual support:** TM can be applied to a variety of languages, which enables the extraction of relevant information from documents written in different languages.
- **Integration with BIM:** TM can be integrated with BIM, which enables the extraction of relevant information to be integrated into the BIM model, providing a more comprehensive and accurate representation of the project.
- **Performance enhancement:** TM can enhance performance by providing insights into the project’s stakeholders' interactions, communication patterns, and information flow. The experiment method is employed to measure the performance.

Overall, the third objective can reveal that TM is a powerful methodology that can be used to enhance the performance of BIM pervasive major construction projects by reducing the manual effort and time spent searching for relevant information, providing more accurate and relevant information, and enabling the integration of that information into BIM.

1.3 Main Contributions of the Research

This research makes several significant contributions to the fields of BIM and TM, particularly focusing on improving BIM practice and performance in large-scale construction projects. These contributions are as follows:

1.3.1 Development of the BIMTAPE Framework

This study introduces the BIMTAPE Framework, an advanced analytical tool that integrates BIM with TM techniques to improve decision-making and performance in major construction projects. The framework allows practitioners to extract actionable insights from unstructured data, such as project reports and communications, filling a gap in traditional BIM approaches that typically handle only structured data. By doing so, BIMTAPE supports a more nuanced understanding of project complexities.

1.3.2 Enhancing the Periodic Table of BIM with the BIM Knowledge Graph

This research addresses a critical gap in the "Periodic Table of BIM" by incorporating the BIM Knowledge Graph (BIMKG). While the Periodic Table of BIM provides a foundational, first-level understanding of key BIM concepts, it doesn’t illustrate how these elements interconnect. The BIMKG addresses this by revealing relationships and connections between different BIM knowledge areas, establishing a second level of knowledge that supports a deeper, more practice-oriented understanding of BIM. This helps professionals to visualise and apply BIM concepts in a way that is both practical and interconnected.

1.3.3 Integrating TM and BIM for Major Capital Projects

Unlike previous studies, which often focus on smaller or more theoretical applications, this research applies TM and Natural Language Processing (NLP) techniques in the real-world context of a major construction project. Through analysing a variety of BIM-related documents, the study demonstrates how TM can uncover patterns and insights that are crucial for managing large-scale projects. This contribution is particularly valuable as it expands the scope of BIM practice to include unstructured data, an area that has remained largely untapped in BIM research.

1.3.4 Validation in a Real-World Major Project

Although this study does not rely on multiple case studies, the BIMTAPE Framework has been validated in a real-time major construction project, as explored in Experiment 2. This single-case approach allows for an in-depth, context-rich examination of the framework's effectiveness within a complex project setting. The methodology behind the single-case selection is well-justified in the research design, demonstrating how BIMTAPE can deliver practical, scalable insights in live projects.

1.3.5 Systematic Approach to Handling Unstructured Data in BIM

The study provides a structured, replicable process for analysing unstructured data within BIM, including data pre-processing, TM model building, and insight generation. This systematic approach addresses a critical gap in conventional BIM practices by enabling the use of textual data that is often overlooked, enhancing the range of information that can be utilised for performance improvement in construction projects.

1.3.6 Guidelines for Industry Professionals

The BIMTAPE Framework offers actionable guidelines for BIM practitioners and project managers. By integrating unstructured data into BIM workflows and helping users establish meaningful connections between knowledge elements, the framework promotes more informed decision-making, collaboration, and efficiency in large-scale project delivery. This is especially valuable in the AEC industry, where a holistic understanding of project information can drive significant improvements in performance and project outcomes.

1.4 Research Methodology

The new technical solution under the research aim is a TM approach to performance enhancement in BIM pervasive major project delivery. The study has three objectives which have been achieved through a combined use of three methods, including Literature Review, Framework (BIMTAPE) Development, and Experimental Case Study. These methods are chosen to work together, by focusing on the retrieval and analysis relevant sets of data and information in a novel way, to achieve the three research objectives, and these include

- The systematic literature review provides an understanding of the current state of BIM and TM in project delivery.
- The BIMTAPE Framework development involves creating a framework for integrating TM into BIM.
- The experimental case studies test the effectiveness of the proposed solution in real-world scenarios.

A further description about the use of these methods is give below.

1.4.1 Systematic Literature Review

The analysis of academic articles reveals the previous direction of the research as well as important concepts that will help in the fulfilment of research objectives. The systematic literature review will be conducted using the theory of inventive problem solving known as TRIZ approach (Gadd, 2011), which is a method used to analyse problems and develop innovative solutions. This approach is used to understand the current state of BIM and TM in project delivery, and to identify potential challenges and opportunities for integrating these technologies. The systematic literature review using the Nine-Square Process (NSP) (Gadd, 2011; Chen et al., 2021; ASQ, 2022) (See Section 2.1) provides a foundation for the research by providing a comprehensive understanding of the current state of BIM and TM in construction or built environment project delivery. It shapes the conceptual framework for the research by identifying the key concepts, theories, and practices related to BIM execution planning and TM in construction or built environment projects. This helps in explaining the purpose and rationale of the research by highlighting the gap in the current state of knowledge and the need for a TM approach to BIM in major construction project delivery. The systematic literature review also helps in identifying the key challenges and opportunities for integrating TM into BIM, and in defining the research objectives and expected research outcomes.

The rationale for conducting the systematic literature review using the NSP approach is that it is an effective method for analysing problems and developing innovative solutions. This method helps the author to think creatively and systematically about the problem, which enhances the conception of the research from the perspective of the author. Additionally, the NSP method is known for providing reliable and valid information by using a structured approach to identify and evaluate the relevant literature on the topic. This can be beneficial in ensuring that the research is well-informed, and that the conceptual framework is built on a solid foundation of existing knowledge. Overall, by using the TRIZ approach in the systematic literature review, the research will have a better chance of achieving its objectives and producing meaningful outcomes that are well-supported by the existing literature on the topic.

1.4.2 BIM Text Analysis

The BIMTAPE Framework was structured using elements from the Royal Institute of British Architects (RIBA) template (RIBA, 2015), and the Cross-Industry Standard Process for Data Mining (CRISP-DM), which provides a structured, step-by-step approach to data mining (Chapman et al., 2000). The RIBA template, recognized as a prevalent methodology for planning and executing construction projects, furnishes a meticulously defined and structured approach to project management. Specifically, the RIBA stages encompass distinct phases for the effective progression of architectural endeavours. CRISP-DM is a widely used methodology for data mining and analytics, which provides a clear and structured approach to the process of understanding, preparing, and modelling data for analysis. Combining the RIBA template and CRISP-DM provides a comprehensive matrix for integrating TM into BIM in a way that is aligned with industry best practices, and that ensures the research objectives are met, relating to the justification of the effectiveness of the new methodology for pervasive major construction project delivery.

1.4.2.1 Incorporating RIBA Template

Incorporating the RIBA template enables the framework to provide a structured approach to project management, which will be useful in evaluating the effectiveness of the new methodology in delivering projects. The well-defined stages of the RIBA template contribute significantly to assessing and enhancing project management aspects within the BIM context.

1.4.2.2 Incorporating CRISP-DM

The incorporation of CRISP-DM into the framework facilitates a structured approach to data analysis. This structured approach proves valuable in scrutinising project processes and understanding their variability stemming from BIM adoption. By using these methods, the research provides evidence of the effectiveness of the new methodology, and identifies areas where the methodology can potentially be improved. Additionally, the research is able to identify any project processes that are affected by BIM adoption and to understand the variability in these processes. Overall, the use of these methods in the development of the BIMTAPE Framework helps achieve the research objectives, and ensures that the study produces meaningful and actionable outcomes.

1.4.2.3 BIM Knowledge Graph Integration

In addition to the BIM Text Analysis and BIMTAPE Framework, the research incorporates the BIMKG. This integration plays a pivotal role in enhancing the overall framework's capability to align with industry standards and improve the understanding and implementation of BIM practices. The BIMKG Integration segment ensures that the research produces meaningful and actionable outcomes, contributing to advancements in the field of BIM and construction project delivery.

1.4.3 Experimental Case Studies

Experimental case studies involve the collection and analysis of data from real-world projects in order to test the proposed methodology and to evaluate its effectiveness in practice. Experimental case studies are useful for evaluating the practicality and applicability of a new methodology, and for identifying any issues that may arise in its implementation. In this research, experimental case studies are used to test the BIMTAPE Framework developed using the RIBA template and CRISP-DM methodologies. The data collected from the case studies is analysed using the BIMTAPE Framework to evaluate the effectiveness of the new methodology in delivering major projects. The results of the case studies provides valuable insights into the practicality and applicability of the new methodology and is used to justify its effectiveness and to make recommendations for future research.

The first experiment is based on data sourced from academic articles, which is a common approach in research as it allows the author to examine existing data on a specific topic, without the need to collect new data. By drawing the first case study from academic articles, the author is able to build on existing knowledge, and to evaluate the proposed methodology in the context of previous research. This can provide valuable insights into the effectiveness of the new methodology. Additionally, using academic articles as a source of data allows the author to examine a wide range of projects and to identify trends and patterns that may not be apparent in a single case study. By using academic articles as the source of data for the first case study, the author is able to gain a comprehensive understanding of the topic, and to evaluate the new methodology in the context of existing knowledge.

The second experiment uses a real-time case scenario by analysing Crossrail project documents with regard to the use of BIM in major construction projects. This approach is useful as it allows the author to examine the use of BIM in a real-world project, and to evaluate the proposed methodology in the context of a current project. By analysing Crossrail project documents, the author is able to gain a detailed understanding of the use of BIM in the project, including the processes and procedures that were used, the challenges that were encountered, and the outcomes that were achieved. This will provide valuable insights into the effectiveness of the new methodology in practice. Additionally, by using the Crossrail project as a case study, the author is able to evaluate the use of BIM in a major construction project, which will provide a useful context for the research and will help to make the findings more relevant to practitioners.

The third experiment is conducted through text collection using the BIMTAPE Framework, which includes regulation documentation, British standards, national policy papers, and enterprise policy papers. This approach will provide a systematic way of collecting and analysing data on the use of BIM in regulatory and policy documents. By using the BIMTAPE Framework, the author is able to identify the key components and principles of BIM adoption in these documents, and to evaluate the proposed methodology in the context of existing regulations and policies. This will provide valuable insights into how BIM is being implemented and governed in practice. Additionally, by collecting and analysing data from a wide range of regulatory and policy documents, the author is able to gain a comprehensive understanding of the use of BIM in these contexts, and to evaluate the new methodology in the context of existing regulations and policies.

In addition to these research objectives oriented descriptions in this section, Chapter 3 provides a comprehensive description about the methodology adopted in this PhD research.

1.5 Research Findings

The research outcomes validate the research aim and objectives, encompassing a detailed grasp of BIM execution planning in major construction projects, introducing the BIMTAPE Framework, and evaluating the proposed TM approach for BIM pervasive major project delivery. Additionally, the outcomes analyse project processes under BIM's influence, assess the methodology's practicality within existing regulations, and offer insights to enhance BIM adoption. These findings inform future research directions and practical BIM implementation in major project delivery.

1.6 Research Implications

1.6.1 Academic Research Implications

The implications of this research resonate significantly within academic circles, potentially pioneering a novel TM approach tailored for BIM pervasive major project delivery. This innovation stands poised to revolutionise the construction industry by enhancing project outcomes and curbing costs. Moreover, the insights garnered from this study have the potential to augment the understanding of BIM execution planning and BIM adoption, contributing to the scholarly body of knowledge surrounding BIM's integration within construction projects. By assessing the proposed TM approach's efficacy against existing regulations and policies, this research bridges the academic-practical divide, offering insights that hold invaluable relevance for both future research directions and the practical implementation of BIM within the major construction projects.

1.6.2 Professional Practice Implications

In the realm of professional practices, the repercussions of this research extend to the realms of policy, regulation, and real-world application. The envisaged TM approach could potentially serve as a ground-breaking tool within the construction industry, offering tangible benefits by refining project outcomes and cost-effectiveness. Furthermore, the insights garnered from this study offer a pragmatic perspective on

BIM adoption enhancement, potentially leading to the optimization of BIM implementation strategies in practice. By evaluating the proposed TM approach against the backdrop of existing regulations and policies, this research equips industry professionals with the knowledge to navigate the intricacies of BIM integration, fostering a more informed decision-making landscape. Overall, the study's implications extend beyond academia, potentially reshaping industry practices through cutting-edge methodologies and insights.

1.7 About This Thesis

This thesis consists of the following seven chapters:

- Chapter 1: Introduction
- Chapter 2: Systematic Literature Review
- Chapter 3: Research Methodology
- Chapter 4: BIMTAPE Framework
- Chapter 5: Justifying New Methodology for Major Project Delivery
- Chapter 6: Findings and Discussion
- Chapter 7: Conclusion

Brief descriptions about these chapters are given below.

Chapter 1: Introduction. In this chapter, the research begins by providing an introduction to the background and motivations that have prompted the study. It specifically highlights the gap analysis, identifying areas within the field that require further exploration. The research scope is defined, outlining the boundaries and contexts within which the study will operate. The chapter also elaborates on the research questions that will guide the investigation. Moving forward, the research aim and objectives are laid out, giving a clear sense of purpose and direction. The chosen research methodology is introduced, with an emphasis on its systematic approach. Key components of the methodology include a systematic literature review, the development of the BIMTAPE Framework, and experimental case studies. The chapter concludes by hinting at the implications that the research findings might have.

Chapter 2: Systematic Literature Review. This chapter commences by providing a comprehensive overview of the research territory. It addresses team performance within major project delivery, outlining measures of team productivity, factors influencing construction team productivity, and the transformative impact of BIM on team dynamics. BIM adoption in major project delivery is discussed, presenting it as a technological and innovative solution to challenges. Stakeholder complexities are explored, reflecting on the challenges and interconnectedness of stakeholders in major construction projects. The chapter proceeds to highlight the learning acquired from major projects and the benefits of BIM in infrastructure projects, emphasizing quality, collaboration, coordination, time, and cost management. It also delves into technical solutions for performance enhancement, digital solutions in construction, and the role of industry guides and standards. The potential of AI and machine learning is examined within the context of the construction sector. The chapter concludes by introducing the theory of BIMTAPE Framework and its applications, followed by a critical appraisal of the literature.

Chapter 3: Research Methodology. This chapter delves deeper into the research methodology by expounding on the significance of TM in enhancing performance in the domain of BIM. It discusses the importance of TM's outcomes, such as increased efficiency, improved quality, and enhanced collaboration. The chapter critically evaluates the options considered and justifies the selected methods. Limitations are discussed, shedding light on the potential challenges that might arise. The research strategy is outlined, encompassing systematic literature reviews, TRIZ-led Evidence-Based Learning (EBL), ontological modelling, and the application of AI in the construction sector. The TM experiment's details are explained, including the processes of data collection, qualitative text analysis, quantitative analysis using SNA, and considerations of validity and reliability.

Chapter 4: BIMTAPE Framework. This chapter embarks on a detailed journey through the experimental phase of the research. It introduces the challenges inherent in text analysis and proceeds to describe the development of the BIMTAPE

Framework. This framework includes components such as BIM text collection, pre-study business understanding, BIM data modelling (TM), and validation. The first experiment involves academic articles (BIM TS 1) and is meticulously outlined, from data collection and transformation to trend analysis and frequency assessment. The second experiment, focusing on technical files from case projects (BIM TS 2), is presented, describing the tools employed, data collection, natural language processing, and topic modelling. Experiments 3-6 involve diverse types of documents, such as regulation documentations, British standards, national and enterprise policy papers. The data collection, TM processes, and resulting insights are detailed for each. BIM model validation techniques, such as social network analysis and network metrics, are explained in the final section of the chapter.

Chapter 5: Justifying New Methodology for Major Project Delivery. In this chapter, the TM outcomes are synthesized to extract meaningful insights. The role of BIM in major construction projects is discussed in relation to enhanced project performance. The involvement of stakeholders is explored, highlighting the complexity and interconnectedness of their roles. The chapter introduces the BIMTAPE Framework and its implications. The contribution and impact of the BIMTAPE Framework are analysed, emphasizing its potential to bridge gaps and enhance project delivery. The application of the TRIZ is introduced, followed by a comprehensive summary of the chapter's content.

Chapter 6: Findings and Discussion. In this chapter, the focus shifts to the outcomes of the experimentation phase. The findings are presented in a structured manner, summarizing the results achieved through the experiments conducted. A summary of these outcomes is provided, and the implications of these findings are discussed. Existing literature is compared to the research findings, highlighting areas of alignment and divergence. The implications of the research for both the industrial and academic domains are explored, suggesting avenues for future research and advancements in the field.

Chapter 7: Conclusion. The final chapter of the research begins by recapping the research's core objectives and assessing how effectively they have been accomplished. An overview of the entire research journey is provided, highlighting the main milestones and contributions. The chapter acknowledges the research's limitations and outlines potential directions for future research endeavours. Tailored recommendations are provided for various stakeholders within the construction and BIM domain, including policy makers, industry professionals, companies, and the general public. The chapter concludes by reiterating the overall conclusion and the significance of the research within the broader context.

1.8 Summary

This chapter sets the stage for this research endeavour by delving into its background, motivations, scope, and key questions. The research aims to leverage TM techniques to enhance performance in the context of BIM pervasive major project delivery. This chapter articulates the research's overarching goals and objectives, outlining a clear roadmap for the exploration ahead.

Through a meticulous approach, the research methodology is outlined, encompassing systematic literature review, BIMTAPE Framework development, and experimental case studies. These methods are strategically selected to extract valuable insights from a diverse range of sources, ensuring a robust foundation for the study.

The chapter also underscores the research's implications and the potential contributions it holds for the field of BIM and project delivery. By addressing gaps in existing knowledge and offering innovative solutions, the research aims to contribute to the enhancement of BIM practices and project outcomes.

As this chapter culminates, it lays the groundwork for the subsequent sections, emphasizing the interconnectedness between the different facets of the research. The upcoming chapters, including the systematic literature review, will build upon the foundation established here, enabling a cohesive and comprehensive exploration of the research objectives.

Chapter 2
Systematic Literature Review

2.0 Background

Prior to the systematic review, some important studies that have focused on BIM and its application in the construction sector will be reviewed. Some of the studies shed light on the BIM's revolution from being a three-dimensional model four, five and now six-dimensional model. Ali Musarat et al. (2022) conducted a systematic review of digitalisation in the construction sector, leading to improving the quality of life. The authors focused on some of the goals in the digital era, including smart and sustainable construction, wireless technologies, economic and architectural growth and BIM's role in attaining the specified goals. The authors confirmed that BIM, along with pre-fabricated construction, is highly effective in enhancing the quality of construction projects. Ali Musarat et al. (2022) confirmed that BIM is a powerful tool in the construction industry, especially for designing and constructing buildings. The author proposed a BIM-based methodology for mitigating carbon emissions and energy efficiency in construction projects. It was also observed that the BIM framework helps to enhance productivity by overcoming challenges (Stojanovska-Georgievska et al., 2022). The potential of BIM in providing monitoring measures was also discussed by Ali Musarat et al. (2022). BIM also manages the information flow and exchanges the data. It is the smart solution to make the construction process sustainable.

The construction industry is notorious for several unsustainable practices that cause substantial loss to the environment and ecological system. Research findings by Ali Musarat et al. (2022) are notable on this background; however, they did not use TM along with BIM. The present research further tries to fill this gap.

Gopal Krishnan et al. (2017) elaborated on the scope of these revolutionary phases (See Table 2.1 below).

Table 2.1: The scope of these revolutionary phases

BIM Revolution	**Functions**
BIM Three-dimensional Model	Serving as a resource for information about a facility, a foundation for decision-making
BIM Four-dimensional model	Construction scheduling information was added.
BIM Five-dimensional Model	The cost information was added.
BIM six-dimensional model	Asset information was added for facility management.

Source: Gopalakrishnan et al. (2017)

Gopalakrishnan et al. (2017) further conducted an in-depth systematic review to explore some research on Big Data issues and apply the TM technique. A three-step is employed for using the TM approach in BIM. The authors suggested an NLP-based technique to process and extract desirable data from the BIM. It is also pointed out that the big data sources.

However, Gopalakrishnan et al. (2017) did not focus primarily on BIM, TM and its desirable impacts in enhancing project performance. However, allied information could be extracted from this study.

Pan et al. (2020) noted that BIM-related technology has become one of the widely discussed topics in the academic body. The study also proved that the innovative text-mining-based model could be developed for BIM patent analysis. The authors highlighted the integration of computer vision technology in BIM to promote rapid development in the construction sector. The authors also confirmed that BIM, IoT, and blockchain could be utilised to prepare a robust building maintenance system. Like Pan et al. (2020), Liu et al. (2019) have also suggested the combination of the BIM framework and blockchain technology for planning and coordinating various stages in the building life cycle. However, BIM has some challenges, especially in legal issues and intellectual property-related matters. This combined approach of BIM and blockchain technology will help to resolve them. Pan et al. (2020) confirmed that BIM increases performance and productivity. Its awareness and strategic application can help to promote innovation in the construction industry. The study also explored the development and new trends in the BIM field. The study also employed the Derwent Innovation Index (DII) database as an effective source for analysing the trend in the BIM field. Pan et al. (2020) also observed that BIM has been a critical tool for its capacity to manage functional and physical digital presentation.

Wen and Xiao (2020) developed a sophisticated BIM model and processing modules. The authors also developed a machine-learning engine embedded with a text processing module and a machine-user interface for visualising the BIM model's data. The objective behind developing the model was to process BIM structured and unstructured model data. Structured data consisted of quantitative factors, including calculation, statistics and geometry. Unstructured data included qualitative components such as text phrases. The machine learning engine was used for the BIM model, making it complex but comprehensive. Wen and Xiao (2020) derived three outputs: [1] object and property retrieve, [2] similar object recognition consistency checking, and [3] completeness accuracy checking. Five key modules of machine learning were discussed in the study (See Table 2.2):

Table 2.2: Machine learning modules

Modules	**Descriptions**	**References**
Text Processing	Extracts textual components such as tokens, words, and phrases, including word-correcting processes.	Wen and Xio, 2020
Named Entity Recognition	Extracts and categorizes target entities for further analysis, aiding in BIM model data classification.	Wen and Xio, 2020
Regexes Pattern Checking	Matches and manipulates textual data using regular expressions for pattern search and substitutions.	Wen and Xio, 2020
Lexical Semantic Relations	Focuses on meaning and context, generating synonyms and establishing relationships among objects.	Liu et al., 2020
Unsupervised Lexicon-Based Checking	Enhances compliance performance by using unsupervised learning to cluster words and ensure regulatory accuracy.	Zhang and El-Gohary, 2017; Berners-Lee, 2016

Three steps of text-based analysis are suggested; they include preprocessing, feature generation, and knowledge mining (Baek et al., 2021). Preprocessing is the initial phase in pipelining the TM process. Text-based research in construction primarily aims at generating new and unique knowledge in the construction and project management sector. Text-based research has also focused on integrating domain knowledge with a variety of state-of-the-art text analysis methods and domain knowledge. Nowadays, construction firms have started developing in-house automated text analysis tools. It is also confirmed that text analysis has become one of the promising fields due to its vast knowledge and NLP and TM techniques, which help to improve the overall performance.

Customised NLP tools and techniques have been developed in various sectors, including construction (Luque et al., 2019; Xu and Cai, 2021). Though it is challenging, a holistic and robust technical ontology can be developed in construction and project management domains. One prominent characteristic observed in the text analysis research is the incorporation of multi-lingual tools. Though most of the language models and text processing tools are based on English langauge, the other languages have also been incorporated in these models. For example, some scholars have developed text-processing tools in Chinese (Wang et al., 2019, Zhong et al., 2019, Jiang et al., 2016), Korean (Choo et al., 2019, Moon et al., 2018., and Kim and Kim, 2018) and French (Kovacevic et al., 2008). Some of the NLP models are language-independent as well (Baek et al., 2021).

According to Zhou et al. (2019), a text-mining approach in the domain knowledge can capture various BIM application needs. Capturing these needs means identifying problems, goals and preferences (Government Digital Services, 2017). It was also observed that there is a huge information asymmetry. To overcome this issue, the developers should develop a BIM app with special attention and effort on the issues, including low sentiments such as reality capture, interoperability, and structural simulation and analysis (Zhou et al., 2019). Conventional and text-mining approaches can capture users' needs. However, the conventional approach has some limitations, and TM is the solution to overcome them (Ahmed et al., 2018). To minimise this gap or limitations, the researchers attempted to extract their needs from the data generated by the users.

The TM approaches have several advantages highlighted in previous studies. For example, Zhou et al. (2019) observed that the text-mining approaches can reach the end-users by manipulating a huge amoung of users' data. Architecture, engineering and construction industry extensively adopts the TM approaches (Marzouk and Enaba, 2019). However, its potential to handle practical issues has been a widely discussed topic among scholars. These practical issues are the accident report analysis and the retrieval of reference cases.

The construction sector is complex, consisting of multiple stakeholders with multiple interests and activities. The sector is not merely associated with constructing buildings but also with infrastructural projects like bridge, road, and tunnel construction. Data mining-based BIM technology can be useful in designing models for these activities. For example, Qu (2021) developed a system or model using BIM and multimedia using various components such as storing construction-related information in the host of the ground command centre, developing a core analysis system to improve the underground shield computer data collection and analysis, and establishing the construction site monitoring systems.

In constructing tunnels and other such infrastructure, the potential benefits of BIM technology and its application have been discussed in several studies (Giani et al., 2020; Moayedi et al., 2019; Stanisz et al., 2019). Qu (2021) conducted research in the context of tunnel-building activities in China. Research in this activity is still in its developing stage in the country. Given this background, his research is worth studying, though it has outlined the application potential of BIM technology in tunnel construction. The author's study is worth studying and reviewing because it focuses primarily on mitigating construction risks due to data insufficiency. The model mentioned in his study is related to tunnel construction risks. However, this model can have the potential to be used in other construction projects as well.

The study by Maksudovna et al. (2023) identified various advantages of incorporating BIM and using TM in the construction sector. Assessing project cost is of utmost importance for ensuring that the construction project does not exceed the budget. They highlighted several advantages of the BIM-based model in managing the cost of the project. TM informed BIM model helps to reduce the labour cost. It also has the potential to form a budget using a database of aggregated unit prices at various phases of construction projects. TM informed BIM model has the capacity to take into account the actual data, collect information about project details and update the Database of large unit prices (Maksudovna et al., 2023). Their study is vital because TM methods can be useful in achieving several construction goals. Though the data mining technique cannot directly extract the knowledge from the text data, it enables to process of text data using TM technology. Researchers have applied TM techniques to extract relevant data from the massive unstructured data. Due to its immense potential, TM-based BIM has gained popularity in technical fields such as architecture, engineering, and project management. In Russia, the use of BIM has become mandatory at all state-ordered facilities in the Russian Federation (Maksudovna et al., 2023).

The present research is similar to the study that Yan et al. (2022) have conducted. Like this research, they also conducted a systematic review of articles on the potential application of TM in the construction sector. TM technique was first introduced in the Architectural, engineering and construction sectors at the beginning of the 21st century. However, the development was slower than in the other sectors (Scherer and Reul, 2000; Chui, 2017). Yan et al. (2022) confirmed that the majority of the data in the construction sector is unstructured, and the TM method is effective in processing this data.

Earlier, it has been discussed how TM is useful and what benefits of BIM TM can be availed to improve the sector’s performance. Lin et al. (2020) also focused on TM’s important role in understanding the on-site inspection of construction projects. It has been observed that 80 percent of the data is in textual format (Martínez-Rojas et al., 2016), and in construction projects, especially, massive data is available in documents such as reports, contracts, policy papers, correspondence, etc. However, mining unstructured data in the construction sector is still in its infancy. Therefore, Lin et al. (2020) have suggested various data mining techniques.

The above studies extended the peripheries of knowledge in BIM and TM in construction projects so that they can avail themselves of their various benefits.

In order to justify the need for this PhD research, a systematic literature review has been conducted with regard to a preliminary review of research background, theoretical considerations, and practical orientations. In the process of this systematic literature review, the following 11 main steps are made, and these include

- Step 1: Strategy for Systematic Literature Review
- Step 2: Team Performance in Major construction Project Delivery
- Step 3: BIM Adoption in Major Construction Project Delivery
- Step 4: Learning Legacies from Major Projects
- Step 5: Technical Solutions for Performance Enhancement in Major Construction Project Delivery
- Step 6: Industry Guides and Standards for BIM Executions
- Step 7: BIM Execution Potential in Major Construction Project Delivery
- Step 8: Applied AI for major construction project delivery
- Step 9: Theoretical Advances in Data, Information and Knowledge (DIK) Applications
- Step 10: The Theory of BIMTAPE Framework for BIM pervasive major project delivery
- Step 11: Critical Appraisal of the Literature

Descriptions about these main steps to form the systematic literature review are given below in the next 11 sections to form this chapter. The interconnections among these steps are illustrated in Figure 2.1 via a Nine-Square Process (NSP) (Gadd, 2011; Chen et al., 2021; ASQ, 2022) in Section 2.1 to form the strategy of this systematic literature review.

2.1 Strategy for Systematic Literature Review

This systematic literature review seeks to address the research gap pertaining to the utilization of TM techniques in conjunction with BIM for performance enhancement in major construction project delivery. By conducting a comprehensive examination of the existing scholarly literature, this study aims to identify the unexplored potential and opportunities for leveraging TM methods to extract valuable insights from textual data associated with large-scale construction projects where BIM is widely adopted. This research endeavours to contribute to the advancement of knowledge by bridging the divide between TM and BIM, ultimately leading to more informed decision-making, improved project outcomes, and enhanced overall project performance.

2.1.1 Overview of Major Construction Projects

Major construction projects are by definition of immense scale in various dimensions, including in terms of physical material, financial investment, and diverse stakeholders. Due to their inherent complexity, they have proved to be intractably difficult to tame systematically in terms of project management, including with regard to long-time frames, enormous physical and logistical scope, and massive expense (and financial risk). Furthermore, major construction projects are additionally difficult to manage (Chen et al., 2021), because they are associated with far-reaching changes to people and the geography of the areas where they are constructed.

Major construction projects that are not managed adequately can have catastrophic ramifications for individuals (particularly clients and investors), local communities, and even the natural or urban landscape. Such projects are identified by their high-cost attributes, which encompass long duration, program urgency, necessitate the involvement of multidisciplinary input, and lead to need for virtual enterprises in their execution. Additionally, even when implemented and managed with normative effectiveness (Chen et al., 2021), they often lead to lower-than-predicted revenue stream and cost overruns.

Aside from the aforementioned attributes, the operational cost is a formidable consideration that is frequently far beyond the initially projected cost, which hampers economic growth and defeats the fundamental purposes of major construction projects. As per Zhang et al.~~,~~ (2022), The Federal Highway Administration (FHWA) defines “major construction projects” as expensive or high-cost, or which attract massive political and public interests because of the massive impact that they tend to have on the environment, the state budget, and the community, typically with a time frame exceeding five years (Youcef et al., 2012). The Major Projects Authority (MPA) provides a comprehensive definition of major projects, classifying them based on specific criteria such as cost, complexity, and significance (Major Projects Authority 2016). According to the MPA's definition, major projects encompass endeavours that exhibit substantial financial investment, intricate operational intricacies, and notable strategic importance. These criteria collectively delineate the scope and magnitude of projects that fall under the designation of "major projects." The MPA's definition serves as a benchmark for identifying projects that warrant specialized attention due to their extensive scale and implications. The complexity associated with the management of major construction projects emanates from the fact that multidisciplinary teams are required to undertake and oversee complex activities, while demonstrating strict conformity to the stipulated deadlines and budgetary allocation.

Given the size of major construction projects, conflicts, poor coordination, and high risks are encountered both in the planning and implementation phases of the project (Santoso et al., 2020). Most project-based organizations possess core competencies necessary for effective and efficient project execution. However, the evaluation of project success is a somewhat difficult task due to the lack of clearly defined criteria for appraising construction project success, and the differing priorities of different stakeholders (Baccarini, 1999). In a bid to understand major construction projects and elucidate areas that ought to be addressed to define and achieve project success, this study seeks to address the existing research gap on project related factors pertaining to major construction projects.

The functionality and communication of the Architecture, Engineering, and Construction (AEC) industry has exhibited substantial changes in recent years because of an extensive adoption of BIM in built environment projects (Moreno et al., 2019). BIM refers to a set of associated processes, policies, and technologies to communicate, produce, and analyse the constructive models such that stakeholders can engage in collaborative design, operation, and building of the facility (Miceli et al., 2019). This study has undertaken a systematic literature review using the NSP (as described previously and in Section 2.1.2 below), in order to justify the need for research into

"A text mining approach to performance enhancement in BIM pervasive major project delivery".

2.1.2 Nine-Square Process (NSP)

TM is the solution of concern to this study, and the review process uses the Nine-Square Process (NSP) tool, as illustrated in Figure 2.1, as adopted from TRIZ (Gadd, 2011; Chen et al., 2021). TRIZ offers a problem-solving tool for forecasting and analysis whose purpose is to offer systematic approach to innovation through the assumption that any given problem can be solved. TRIZ is a unique feature because of the potential problems that are faced aligned to the contradictions that emerge for the situations, which are different that need a common implementation (ASQ, 2022). The execution of TRIZ involves three steps of definition of the problem and expression of the contradiction in a simple and clear manner, approaching the problem as the system and monitoring it, and determination of the powerful principles applicable for the respective problem.

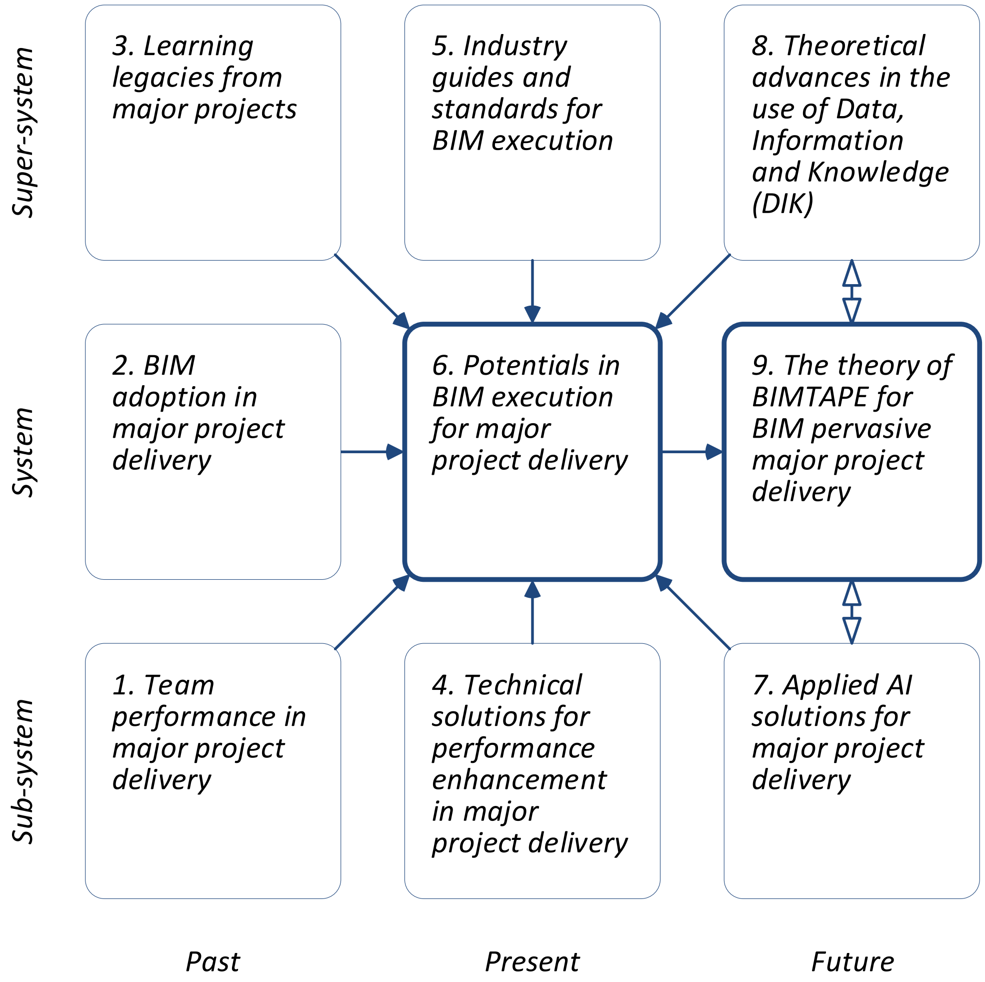


Figure 2.1: NSP for systematic literature review

Figure 2.1 depicts a dedicated NSP, which is adopted for the systematic literature review for this study. The NSP can justify the rationale for research on BIMTAPE Framework. Figure 2.1 identifies the nine items that are taken into consideration when conducting the systematic literature review. The NSP depicts the technique for the exploration of the issues alongside their impacts through the examination and evaluation of their present, past, and future (ASQ, 2022). The adoption of this technique is based on the need to ascertain that companies have the chance to evolve and change without having to rely on the past tactics that may fail to work efficiently within the current world. The nine windows ensure that it is possible to separate the problem in time and space.

2.2 Team Performance in Major Project Delivery

Major construction project delivery requires the establishment of a team, which *ipso facto* entails team building. Poor team building results in project failure due to inherent role conflicts, different interests, priorities, judgments, and outlooks, among the team members. The team has to be built with a clear and concise objective, which steers it towards the realization of the set goals and objectives (Baiden and Price, 2011). Leadership is also vital in the team, where structures and roles and responsibilities of the team members are outlined to achieve success for the team. The team leader has to handle conflicts, encourage commitment among members, address communication problems, and motivate group synergy through clarification of the anticipated objectives for a given project (Manata et al., 2022). The process results in ensuring that it is possible to achieve good performance for the team in major construction project delivery.

Project team integration in construction involves the collaboration of the behaviours, working practices, and methods of stakeholders, particularly supporting the free exchange of information among parties (Ghosh et al., 2020). An integrated team environment depicts the sharing of knowledge and skills with the aim of improving the delivery of the project (Mariam et al., 2022). The integration of the project team also contributes to the improvement of project team performance in the delivery of the major project. The situation occurs since teamwork is regarded to be one of the traits of the construction industry (Assaf and Hassannain, 2014).

The delivery of construction projects is achieved by different professionals who work together to realize a specific goal. Research indicates that knowledgeable leadership assists in promoting the attainment of project success by convincing individuals the need for the change and encouraging them to work together in complex environments to achieve project goals (Assaf and Hassannain, 2014). BIM is a recent technology, which PMs utilize as a means of improving coordination, collaboration, and communication to support the attainment of team performance in the delivery of the projects.

Operational efficiencies for the project teams are improved through the adoption of BIM Collaboration Format (BCF), a communication standard that facilitates the exchange of issue-related information within BIM models (Eisenmann and Park, 2012). The situation occurs since BIM affect the approach that is adopted by professionals in different teams when working on projects (Hou and Wu, 2019). BIM creates an opportunity for the professionals to engage in early stages of the project, where they support innovative design solutions and solve conflicts for the building prior to the onset of the conflicts for the structure (Wang et al., 2016). Team performance is also enhanced through BIM, since it is possible to identify obstacles, risks and challenges for the projects and solve them without having to rely on obstacles that are costly. These benefits contribute towards the attainment of excellent delivery for the major project.

However, the AEC industry has not yet fully adopted BIM and associated data-mining approaches in many contexts (Zhou et al., 2020). TM approaches are adopted for information extraction or retrieval, recognition of terms, keywords or entities, topic models, sentiment analysis, and document classification or clustering. Deterministic or probabilistic approaches are adopted in TM (Raouf and Al-Ghamdi, 2019). The AEC industry has TM applications, which are recorded at infancy stages. As such, TM needs to be adopted to ensure that there is performance enhancement for the major construction project delivery (Shmueli and Koppius, 2011). The approach helps in recording and documentation some limitations on BIM adoption and implementation that include quality effects and time and persistent cost overruns.

2.2.1 Measuring Team Productivity

The construction industry has given rise to distinct perspectives on the constitution of measuring team productivity, which is reflected in visible differences in the ways in which data collection and analysis are undertaken, as well as the quality and scale of data. According to Dozzi and AbouRizk (1993), the team measurement viewpoints can be conceptualized as macro and micro perspectives. Inconsistencies in the measurements and the lack of consensus on the constituents of measurement (Allmon et al., 2000; Goodrum and Haas, 2002) are fundamental causes of controversy in terms of macro perspective measurement methods. The macro perspective gives room for identifying the trends in the long run visible at the industry level (Teicholz, 2004 and 2013). Several scholars have discovered that multifactor productivity (Tran and

Tookey, 2011), which is a combination of team and capital costs producing output, is best suited for the macro perspective. However, the usefulness of macro perspective in organizations is questionable, since it does not provide a baseline for consistent data comparison and the interpretation might be misleading (Goodrum and Haas, 2002).

Dozzi and AbouRizk (1993) described the micro perspective as being task-oriented, whereby it is generally preferred by organizations due to its amenability to tangible outcome benchmarks for team productivity. Using this perspective, productivity is gauged on the basis of the input and output of the tasks assigned to individuals (Halligan et al., 1994; Park et al., 2005). Operational productivity is defined by Durdyev and Mbachu (2011) as the quantity of output from a process per unit resource of input. For instance, output could be classified in units or dollar power with the relation to the manpower used, machinery, materials, or money. With this in mind, *construction* productivity is considered to imply *team* productivity, which Halligan et al. (1994, p. 48) stated has several advantages. Most significantly, this approach explains team productivity in comprehensible and operational ways, with team productivity taken as the most significant source of distinction in construction productivity, while the latter is viewed as the datum for all team productivity measures. Due to the lack of a standard definition of productivity in the construction industry, Park et al. (2005), decided to use the Equation 1 below to define team productivity:

Equation 1: $Productivity = \frac{Output}{Resource\ Input}$

Halligan et al. (1994) insisted that construction productivity needs to indicate the units or work placed or output per man-hour, as illustrated in the Equation 2 below, which is commonly referred to as the unit rate:

Equation 2: $Team\ Productivity\ (unit\ rate) = \frac{Output\ (Quantity)}{Input\ (Work\ Hour)}$

Team productivity can also be measured using the team productivity factor is shown in the Equation 3 below:

Equation 3: $Performance\ Factor = \frac{Estimated\ Unit\ Rate}{Actual\ Unit\ Rate}$

Relatedly, the performance ratio can also be used to measure team productivity, which researchers such as Yi and Chan (2014) and Završki (1999) calculated with no baseline of interruption, as illustrated in the Equation 4 below:

Equation 4: $$Perfomance\ Ratio = \frac{Actual\ Productivity}{Expected\ Productivity}$$

In the field there are many different approaches that can be used to measure and evaluate team productivity, and the most significant dimensions to consider for effective management are comparison between systems and across time. The relativity in the concept of productivity needs to be conceptualized to provide the evaluability indications of performance (Bernolak, 1997). The negative impacts of productivity can be identified and mitigated using methods such as work sampling, field rating, field surveys, five-minute rating and using models such as the construction productivity management model, productivity delay model, and or factor-based models (Thomas et al., 1990; Dozzi and AbouRizk, 1993; Motwani et al., 1995). Models for the quantifying and evaluating the variability of production include baseline productivity analysis and measure mile analysis, which used to determine the usefulness of a project. These act as the indicators of the overall performance, as well as the justification for loss claims for lost productivity (Ibbs et al., 2007; Menches and Hanna, 2006; Schwartzkopf, 1995; Thomas and Završki, 1999).

2.2.2 Factors Affecting Construction Team Productivity

Research on construction team productivity has identified several key factors that influence performance at various levels. According to Yi and Chan (2014), while construction productivity and team productivity are often used interchangeably, there is a distinction in the factors affecting each. Team construction productivity is considered a subset of general construction productivity. Numerous studies (See Table 2.3) have proposed different ways to categorize these factors, highlighting the complexity of the issue.

For example, Kazaz et al. (2008) identified four main categories affecting team productivity: economic, socio-psychological, organizational, and physical, with 36 underlying factors. Rojas and Aramvareekul (2003) developed a framework consisting of four classes manpower, industry environment, external conditions, and management systems with 18 subcategories. In contrast, Dozzi and AbouRizk (1993) identified nine categories with 44 factors, while Enshassi et al. (2007) proposed 10 categories with 45 underlying factors. More recently, Rivas et al. (2011) expanded the categorization to 12 classes with 83 factors.

Other researchers have explored the role of motivation, which has been classified both as a category with sub-factors (Dozzi and AbouRizk, 1993; Enshassi et al., 2007) and as an underlying factor (Adrian, 1995; Dai et al., 2009). Despite these differences in categorization, there is general consensus on the major factors influencing construction team productivity, which are often linked to the industry, organizational environment, or project level (Yi and Chan, 2014; Rivas et al., 2011). At the individual level, factors such as skills, management, supervision, and labour are also significant.

In terms of improving team productivity, several strategies have been suggested, including the introduction of new techniques, technology, and management systems (Jergeas, 2009). However, Rojas and Aramvareekul (2003) argue that improving productivity is challenging and requires not only the introduction of new methods but also the integration of managerial and technological innovations. Barima (2017) emphasizes that technology has a particularly strong impact on productivity, though its benefits may be undermined if the costs of implementing new technology outweigh the resulting productivity gains.

Ultimately, the implementation of new technologies like BIM has the potential to enhance team productivity, provided it is designed, implemented, and maintained effectively. However, if the cost of BIM implementation exceeds the benefits, the resulting productivity improvements may be minimal.

Table 2.3: Categorisation of Factors Affecting Construction Team Productivity

Study	**Category**	**Number of Factors**
Kazaz et al. (2008)	Economic, Socio-psychological, Organizational, Physical	36
Rojas and Aramvareekul (2003)	Manpower, Industry Environment, External Conditions, Management Systems	18
Dozzi and AbouRizk (1993)	9 categories	44
Enshassi et al. (2007)	10 categories	45
Rivas et al. (2011)	12 classes	83

2.2.3 The Impact of BIM on Team Productivity

Although there is relatively limited evidence in terms of empirical data for the impact of BIM, studies have been conducted to quantify the impacts, and evidence has emerged affirming the successes of the BIM and its positive impacts on various project performance dimensions. For instance, Bryde et al. (2013) surveyed 35 projects operating on BIM and reported that most of the projects reported the benefits of BIM related to the reduction in the costs as well as time management. The respondents also pointed at BIM helping PMs in the communication and coordination improvement. However, the research did not consider the impact of BIM on productivity *per se*, which is the foundation of adoption and improvements.

Researchers such as Giel and Issa (2013) focused on the return on investment (ROI) of BIM, which includes conflict detection, leading to avoidance of extra costs. This implies that productivity will improve since there will be less work, as well as design changes. However, they did not discuss the subject in this direction. Other studies such as Suermann (2009) and Cox et al. (2003) tried to describe what organizations need to undertake in the measurement and evaluating the effects of BIM on the performance of projects. In a study conducted by Khanzode et al. (2008), an investigation was undertaken to explore the benefits and challenges of BIM for mechanical, electrical, and plumbing (MEP) stakeholders through coordinating the project at Camino Medical Healthcare facility. The results by this research show that BIM brought about several impacts, including improved understanding of the project by stakeholders, increased field productivity (from 5 to 25 percent), more timely completion, off-site quantity prefabrication, and a reduction of rework amount (to 0.2 percent of overall field work).

Sacks and Barak (2008) tried to quantify the impact of BIM on productivity for structural engineering projects, and concluded that an estimated increase in productivity of 15 to 41 percent is possible, due largely to less time spent on drawing production. An action research project by Coats et al. (2010) proposed several key performance indices, using a small design firm in the UK to study BIM process implementation, which noted increased productivity in relation to man hours per project and speed of completion. A dissertation by Chelson (2010) found that instrumental KPIs in BIM projects included amount of rework, quantity of request for information (RFI), change of plans due to conflict, and schedule compliance.

2.2.4 Exploring the Influence of BIM on Team Productivity

As discussed in the preceding section, BIM increases the prospect of project success even in the context of major projects that have huge budgetary allocation and longer durations than most conventional projects. A broad body of research has evidenced that BIM brings a plethora of advantages when it is optimally integrated with the organization. In this regard, the contribution of BIM towards project success is irrefutable. Nevertheless, Poirier et al. (2015) suggested that since empirical data evidencing the advantages of BIM in the context of mega construction projects is hard to attain, the measurement of this inherent advantages is yet to be standardized.

An empirical study conducted by Alkhiami and Zangana (2017) determined that KPIs can be used in the quantification of the current performance of BIM in major construction project management. In a bid to measure the collaboration between different stakeholders, the researchers used Macleamy Time/Effort curve. The findings supported the assumption that BIM optimizes the management of major construction projects. To be precise, the integration of BIM reduced the need for paper documents, trust and collaboration was enhanced; particularly, BIM fostered closer collaboration between the client and the contractor.

Despite positive correlation between project success and the integration of BIM, Alkhiami and Zangana (2017) determined that the adoption of BIM also resulted in delays, shortage in employee, cost overruns, and psychological impediments which can potentially threaten the implementation of the project. Aside from the usage of KPIs in measuring the advantages of BIM, they also confirmed indicators can be used to determine the effect of BIM on productivity in the context of major construction projects. Figure 2.2 presents the KPIs used by Alkhiami and Zangana (2017) to appraise the impact of the integration of BIM on productivity.

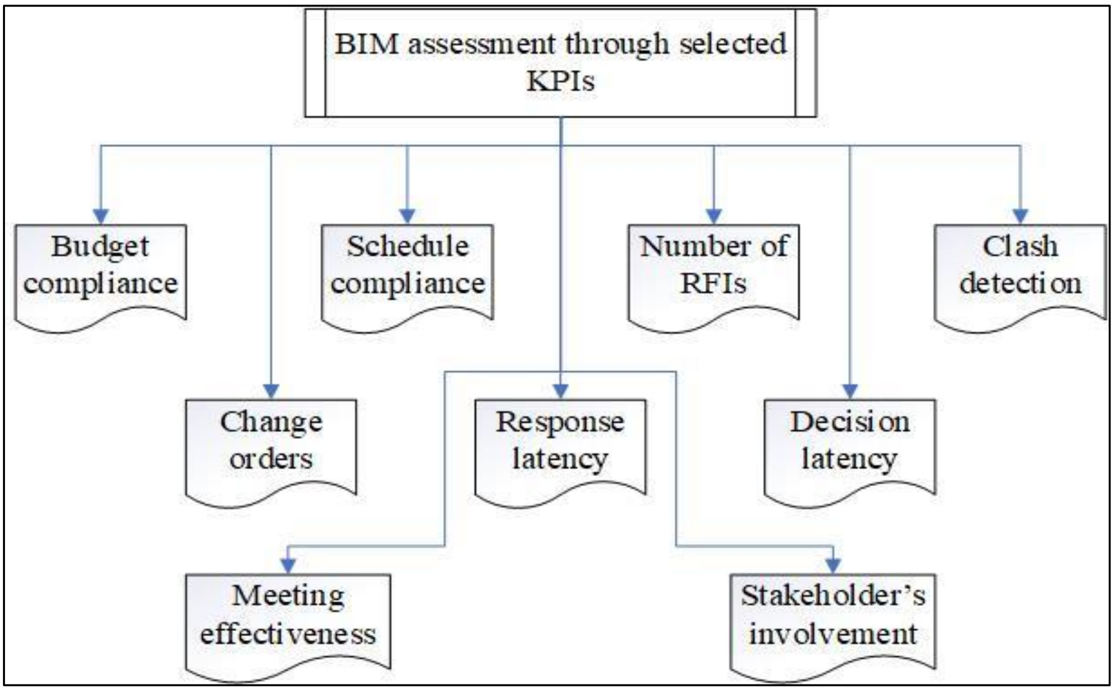


Figure 2.2: The impact of the integration of BIM on productivity (Alkhiami and Zangana, 2017)

Clearly delineated KPIs are fundamentally necessary for organizational performance analysis, and various indicators can be used in construction, including pre-construction man-hours, request of information pertaining to change orders in the course of the construction, issue resolution, the occurrence and incidence of injuries, and variance in the design (Fischer, 2013). The occurrence of these events can be used to measure the major construction project performance. Fischer's (2013) position pertaining to the reduction of RFI was complemented by a study conducted by Chelson (2010), which reported that companies that have used BIM in the course of project management in the past posted a 40% productivity increase. This surge was attributed to the capacity of BIM to squeeze the design phase. Furthermore, the implementation of 3D models as well as the recruitment of competent personnel to oversee the integration of BIM propagates ease in conflict discovery and clashes and other problems that are likely to surface during the design phase, resulting in net savings for the client.

Other studies discussed in the preceding section support the assumption that BIM improves performance and productivity; there are limited examples of how BIM improves performance specifically in the context of major construction projects. As a consequence, major construction projects continue to suffer from budget overruns, loose adherence to set schedules, and quality issues even if an appreciable level of BIM integration is demonstrated.

After focusing on team performance on major construction project delivery, it is clear that BIM has a role to play in eliminating inefficiencies for the teamwork, addressing challenges and risks for the project before they occur, and supporting professional team interaction and collaboration to deliver the project as per the set goals and objectives. Hence, BIM needs to be adopted in major construction project delivery to realize positive outcomes and achieve the set goals and objectives.

2.3 BIM Adoption in Major Construction Project Delivery

As discussed previously, major construction projects face more serious and distinct challenges to the hurdles associated with conventional projects. Consequentially, exploring BIM in the context of major construction projects is integral in exposing how the technology relates to this particular area of construction. The adoption of BIM in major construction project delivery involves business process reforms in individual organizations and industry-wide practices. BIM has been adopted in the construction industry since the advent of the millennium, with adoption focused on increasing efficiency, transforming capacity, increasing project stakeholders' collaboration, and improving productivity (Ullah et al., 2019). BIM adoption creates the opportunity for design visualization, examination of the model reliability, fast alternative design creation, building performance forecasting, and report production (Basu, 1991). However, the implementation rate for BIM has been tied to several barriers, which have been recorded across the globe.

BIM has been pioneered in the USA, China, the UK, and Scandinavian countries (Ullah et al., 2019). However, BIM adoption and implementation has been tentative and even lacklustre in all contexts, because of associated costs for implementation, deficiency in capital, high risk levels for the technology, and the unwillingness to initiate new workflows (Binbin and Xiaoyan, 2017). Furthermore, there are issues of a lack of interest (or trust) among clients, inadequate training, insufficient expertise level, matters of data ownership, and unavailability of standardized protocols and tools (Ullah et al., 2019).

Regions that have already adopted and implemented BIM have significantly depicted the potential benefits that include increase in performance of the project works and delivering high quality projects (Parker et al., 2017). The situation occurs bearing in mind that modern construction projects are more highly complicated than traditional major project, with a commensurate requirement for increased coordination of multiple stakeholders (Miceli et al., 2019). Public adoption of BIM focuses on model management actions, public governance actions, and product management actions, which are integral towards supporting the major delivery of a given project. Furthermore, BIM acts as an efficient tool for planning and scheduling project activities, which leads to better delivery of the major project related to interrelated 3D elements with different attributes and parameters (Nawaz et al., 2021). Thus, successful implementation of BIM in the construction sector calls for revamping of the construction practices that are in existence, whereby the approach can lead to the attainment of efficient results (Olanrewaju et al., 2021).

According to Bawomo et al. (2021), it is undisputable that BIM has been integrated into several areas of the construction industry, particularly building projects. Nonetheless, many organizations in the industry still struggle when it comes to the implementation of BIM in major construction projects such as railway and roads infrastructure projects (Shaignmoghadan and Motamedi, 2021). Organizations that have successfully adopted BIM in major construction projects have reported an increase in time efficiency and the prevention of cost overruns. Also, the adoption of BIM has demonstrated a positive correlation to a decrease in project risks, such as cost surges, the occurrence of construction defects, and delays in the presentation of project deliverables.

Nonetheless, the complexity of major construction projects and the range of professionals required in the implementation of BIM, still presents an impediment to the integration of this technology. The challenge spans from the design to the construction phase of the project (Retzlaff, 2018). This necessitates the exploration of how BIM can be implemented especially when dealing with projects that have long deadlines and encompass huge financial commitment.

2.3.1 BIM as Technological and Innovative Solution

The whole purpose of exploring technological or innovative solutions to lifelong issues in construction is to improve project efficiency and to promote the delivery of major projects. Also, the integration of technological approaches to the management of mega construction projects is intended to reduce the incidence of mistakes and risks that may result in delays as well as an increase in the cost to be incurred by the client. Therefore, in spite of the difficulty associated with the implementation of technological solution, such innovative approaches are still crucial in ensuring that persistent underperformance, especially in the management of major construction projects, is addressed accordingly (Chen et al., 2020).

In this regard, Chen et al. (2020) assert that innovative solutions should be capable of confronting a plethora of issues encountered by PMs in the implementation of major construction projects. The level of adoption of digital solutions is based on how well the approach improves project management processes and address one or more of the aforementioned issues in the management of major construction projects. Technological tools such as BIM have the potential to transform the entirety of the construction process as it changes basic process into more procedure that contributes to optimal construction project management.

Construction technology firms invested a cumulative sum of USD 10 billion from 2011 to the dawn of 2017 to drive research and development aimed at stimulating the development of technological solutions that can be used to improve processes that range from yard inspection and portfolio management to off-site fabrication (Dahlqvist et al., 2018). The development and ultimate implementation of such innovative solutions alongside tried-and-tested tools (such as BIM) extensively promote cost and deadline management while fostering optimal collaboration among involved stakeholders.

As a means to improving operations in the management of mega construction projects, Blanco and Chen (2014), and Kim et al. (2017), suggest that more emphasis should be placed on the development of innovation and conducting research that seeks to determine the major impact of BIM on project delivery. Although different construction projects face different project implementation issues, progress made in the adoption of technology makes alleviation of these challenges possible. For instance, the adoption of 5D BIM, which is a combination of design, scheduling data, and 3D physical models of buildings has presented a positive result as far as project execution is concern. The model can be used in the design phase and simulation of the entire projects with the aim of identifying and addressing looming project risks. The results of the simulations allow construction PMs to make informed decisions on deadlines, budgetary allocations, and the ability to address issues before they arise and hamper project progression.

According to a study conducted by Brent et al. (2018) with the aim of debunking how the full potential of BIM adoption can be realized in the construction settings, the recruitment of persons that possess the necessary core competencies to oversee the adoption of BIM and the overall implementation of the project will allow engineers and stakeholders in the construction project to leverage BIM tools in major construction projects to increase implementation success. The study also determined that the incorporation of government agencies, contractors, and other stakeholders in the implementation of the major construction projects improves the possibility of success.

2.3.2 The Awareness of BIM Adoption by Stakeholders

In a bid to realize the full potential of BIM adoption, stakeholders and construction players involved in major construction projects ought to instigate an organization-wide cultural and paradigm changed. Partial changes can result in partial realization of BIM benefits; as a result, total commitment to this innovation as well as improved reliance on information is core to successfully adoption of BIM in major construction projects. In this regard, recruiting new talent at the highest managerial cadres and the front lines is paramount. According to Sheldon and Shephard (2018), more can be squeezed out of BIM if the following considerations are followed.

The first consideration is the creation and the implementation of a top-down culture that is dependent on data or rather data-centric workflows. This implies that all process with the organization must be informed by collected and synthesized data; this way optimal performance across the organization can be attained. Secondly, the organization as well as its construction players must be prepared to embrace the result attained from the implementation of the data-driven practices. This might mean doing things that may seem orthodox to more conservative employees. The last consideration is pushing data into the field. Availing data to front line personnel will be crucial because it will be equipping them with necessary information that could allow them to avoid the occurrence of cost and time overruns in the course of project implementation (Sheldon and Shephard, 2018).

There seems to be a growing consensus among researchers that the most integral step in the adoption of technology in general (not just BIM) in construction organizations is the introduction of interventions that are aimed at educating construction players about the technology. According to Asvadurov et al. (2018), the integration of technological approaches improves the overall productivity of the workforce as well as the performance of the organization. However, success is guaranteed by how well the team collaborates, enlighten the workforce on how BIM and other related technological tools can be used to propagate collaboration will be integral in ensuring that the personnel can take advantage of such tools to improve operational efficiency.

Asvadurov et al. (2018) also state that project teams must focus four main practices as a means to improve successful management of major construction projects: (1) investing in project teams, including education on relevant innovative solutions such as BIM; (2) promoting prompt decision-making by supplying decision makers with crucial information; (3) the adoption of forward-looking performance management; and (4) team leaders must drive desired behaviours consistently by developing rewarding systems to fasten the transition. By adopting these four practices, the chances of adopting BIM and other innovative solutions in major construction projects can be markedly increased.

The related awareness of BIM and attitudes towards it has improved over the past eight years, an NBS study on the use of BIM determined, according to the National Building Specification [NBS] (2012–2019), which demystified that the growth in knowledge and use of BIM has increased markedly. Concurrently, the construction industry has evolved over the decades as it embraces digital ways of doing things. Nearly all respondents involved in the study were well informed about BIM: only 2% said they were unaware of it, and less than 1% was not sure about it. This collection of NBS annual surveys has been sponsored by BIM-oriented organizations that include the British Standards Institution and the Landscape Institute.

The breadth and depth of implementing BIM systems in construction activities around the world have been suggested by Chen et al. (2020), who used LinkedIn to recognize a steady rise in BIM consciousness through the findings of writers and membership of the BIM network of experts. They hypothesized that the widespread adoption of BIM could transform the construction industry, since it is utilized in the collection of data to performing data-driven activities and process that foster the delivery of major construction projects. Variations in the adoption of BIM exist across the industry.

In their worldwide survey, Young et al. (2008) found that 90 percent of respondents in the field of architecture and 85 percent of all respondents thought of themselves to be fully competent in BIM; while construction contractors declared themselves to be the most experienced BIM users, 46 percent of respondents designated themselves as skilled users. Architects, however, were found to be the most interested parties in the adoption of BIM, while building contractors and PMs were more likely to see BIM as a data management system, while facility managers usually viewed BIM as a system of project information (Gu and London, 2010).

Based on studies conducted by NBS in the UK by Chen et al. (2020), Bernstein et al. (2014), and Jones and Laquidara-Carr (2017), it is notable that skills required to effectively implement BIM, perceived benefits, and BIM experience were limited among players in major construction projects. Complementary studies conducted by Kiziltas and Akinci (2010), Lee and Yu (2016), and Leite et al. (2011), among others, elucidated that there is a lack of specific research to detect awareness and adoption of BIM among stakeholders in the major construction project context.

2.3.3 Stakeholders in Major Construction Projects

Major construction projects involve a wide variety of stakeholders whose requirements and priorities must be considered to ensure the success of the project (Aaltonen, 2011; Patanakul et al., 2016; Teo, 2013). Stakeholders may be persons, for-profit or non-profit entities, or organizations that may influence the project or be influenced by it. According to a study conducted by Aaltonen et al. (2010), stakeholders can be divided into internal and external stakeholders. Typically, internal stakeholders are a structured amalgam or group of project participants who naturally endorse the project. Examples of internal stakeholders are the client and involved government agencies or entities. Conversely, external stakeholders are not formal entities that may be impacted directly or indirectly by the project (Awakul and Ogunlana, 2002). For instance, when constructing a railway station, which is major construction project, the local populace will be affected since an alternative means of transport has been made available.

Major construction projects and their volatility and high complexity call for a comprehensive approach to the management of competing stakeholder needs and relationships, which play a key role in the success of a project. The effective delivery of major construction projects is inherently difficult, and managing stakeholders' expectations and relationships is a prerequisite essential for management. Mok and Shen (2016) suggested that a network-based approach to stakeholder model to investigate both key players and their interests can increase the accuracy of stakeholder management practice in construction, as well as enhance completeness and effectiveness. On the other hand, a new value prioritization strategy was developed by Chen et al. (2020) to define key benefits of BIM in the delivery of both megaprojects and major construction projects. They developed a compass tool to visualize the status of value attainment among project stakeholders who deal with major challenges in the delivery of major construction projects.

Blaine et al. (2014) also affirmed that the cooperation of stakeholders is the most integral factor that fosters the attainment of a successful execution of BIM projects. For example, through the exchange of common knowledge using BIM, the communication process and cross-multidisciplinary collaboration could potentially minimize errors. This postulation was endorsed by Mounir et al. (2018), who stressed that using the collaborative approach is one of the most significant factors in the construction management process. Empirical research by Sahar et al. (2017) elucidated that through cooperation, efficient and effective construction project delivery can be achieved. Furthermore, during the partnership between multi-disciplinary teams, the errors of BIM-based projects can be minimized (Mehrdad et al., 2018). Numerous studies (Prahalad and Hamel, 1997; Thompson and Miner, 2006) have stressed that information sharing is the basis of cooperation between stakeholders in construction projects. Clearly, it is important to carefully study the relation between stakeholders and their effects.

2.3.4 Complexities of Stakeholders in Major Construction Projects

Stakeholders are essentially a groups or individuals who may influence the results of the project, whose living conditions are influenced by the project positively or negatively, and who derive direct and indirect benefits and/or losses due to the implementation of the project (Li et al., 2012). Three facets of the complexity of stakeholders can be studied in major construction projects: The first facet is stakeholder problems and their interconnectedness, referring to what stakeholders are worried with in the project and how these worries or aspects are interrelated. The second facet is stakeholder relationships and interactions, referring to the social connections between stakeholders, for instance, individual and government. The last facet is stakeholder and problem dynamics, relating to how the stakeholder group and the interests of stakeholders shift over time as the project progresses.

2.3.5 Stakeholders Problems and Their Interconnectedness

Major construction projects' growth can affect the economic interests of different groups of stakeholders. Due to the diverse stakeholder perspectives in the evolving project environments, stakeholder issues, which are defined as the vested interests of project players, are often dissimilar and complex. In response to the changing climate, new stakeholders and dissimilar concerns often arise; problem priorities can also differ across various project player groups. According to Li et al. (2012), if they are inadequately catered, the competing stakeholder interests can result in project threats and failures. The same researchers identified major stakeholder issues and how they can be prioritized.

Zeng et al. (2015) elucidated the principal stakeholder concerns related to the fulfilment of major construction projects social responsibility. The prevailing knowledge of stakeholder interests in major construction projects has been reinforced by existing publications. The estimation and prioritization of the magnitude of the problem, nevertheless, leaned entirely on the personal judgment of stakeholder groups; the actual interdependencies between stakeholder issues and the perpetuating effects created by the problem network were not considered. As such, a comprehensive procedure is necessary to examine interdependencies between stakeholders and evaluate their effects on one another.

2.3.6 Interactions and Relationships of Stakeholders

In the context of major construction projects, stakeholders are directly or indirectly linked by several forms of associations across level of the organization boundaries, so rather than being fragmented in vacuum, they are rooted in different social networks. The early body of research paid a lot of attention to structured stakeholder relationships. For example, the contractual agreements between project entities regarding the resource sharing and the provision of construction services and the hierarchical relationships between participants in intra-organizational projects (Pryke, 2004). Recent studies move the emphasis towards informal stakeholder relationships and make substantial efforts to strengthen relationship management as a means to foster project success and optimal management of allocated resources (Chen et al., 2020).

Informal stakeholder relationships were categorized by Cross and Parker (2004) studies into four types: influence, collaborative relationships, knowledge sharing, and interpersonal relationships. According to a complementary study conducted by Chinowsky et al. (2008), communication and knowledge sharing ought to be evaluated in a bid to elucidate how they contribute success in major construction projects. In a project climate, stakeholders do not even exist individually; they are defined by these relationships.

The shift towards BIM adoption in major construction project delivery underscores the necessity of transforming legacy construction practices. This shift aims to promote collaborative efforts among major project stakeholders, providing opportunities for advanced project planning and scheduling. These adaptations contribute to the extraction of valuable learning insights from major projects, thereby enhancing industry practices.

2.4 Learning Legacies from Major Projects

Major projects that have been executed by AEC companies in the UK, USA, Germany, and France have indicated that the use of BIM in projects has positive impact of increasing the productivity and improving the quality of work (Oktem et al., 2020). The completion of these projects indicate that BIM has the benefits of better quality, increased collaboration, time and cost savings, and increased stakeholder coordination (Lawson, 2006). TM for BIM utilization in major projects indicates that BIM assists in tasks that include clash detection, 3D visualization, feasibility studies, visualized scheduling (4D) management, model-based quantity estimation and take-off, creation of the facility management-built model, Leadership in Energy and Environmental Design (LEED) certification or environmental analysis, geospatial and visual coordination, creation of shop drawings, and visualized constructability review (Ali et al., 2022). An example of this was the completion of 2012 London Olympic 6,000-seater Velodrome cycle track. Table 2.4 below provides a good example of case studies on major projects completed through utilization of BIM.

Table 2.4: Crossrail case studies on major projects using BIM

Title	References
Case study	
Young Crossrail Programme	Hillier, 2014
Addressing Skills Gaps Through Direct Intervention	Georgina and Nathan, 2016
Noise and Vibration Controls for the TBM and Temporary Construction Railway	Bird et al., 2016
Information Handover Principles	MacDonald, 2016
Building a spatial infrastructure for Crossrail	Irwin and Tamash, 2016
Vibration Management and Listed Buildings	Bird et al., 2016
Crossrail LFB Emergency Services Liaison	Dave et al., 2017
Works Information Study (Supply Chain Quality Requirements)	Elliot, 2016
Good practice document	
Crossrail BIM Principles	Crossrail Ltd, 2017-2022
Heritage deed for works affecting listed buildings - template	
Heritage deed for settlement mitigation works affecting listed buildings - template	
Crossrail Operations and Maintenance Information Guide	
Crossrail Asset Information Guide	
Journal article	
The Geo-centric Railway - Why Location Matters in the Rail Industry	Irwin, 2016
Crossrail: Building a Virtual Version of London’s Elizabeth line - The Development of the BIM Environment	Malcolm, 2017
Benefits of Contractor Prototyping: GFRC Cladding on Crossrail	Jim and Jorrin, 2020
Micro-report	
Crossrail’s Information System	Catherine and Alistair, 2018
Corporate Information Handover Project - Enabling Data Migration	Tahir, 2018
BREEAM for Underground Stations	Mike, 2018
BIM Metrics	Tahir, 2017
Technical paper	
Ground Settlement Behaviour in Chalk Due to TBM Excavations	Cheng and Mikulski, 2014
The Protection of the 400kV Cables at Pudding Mill Lane	Evans, 2014
Lindsey Street Bridge - A Structural Solution to Settlement Mitigation	McCarthy, 2016
Design and construction of Fisher Street crossover cavern on Crossrail contract C300/C410, London	St. John et al., 2016

Table 2.4: Crossrail case studies on major projects using BIM

Title	References
3D geological model of the completed Farringdon underground railway station	Angelos et al, 2016
Design of the Deep Cut and Cover Crossrail Paddington Station Using Finite Element Method	Mitesh and Aliki, 2016
Covered Way 126: Safeguarding a Brittle Structure over a Live Railway during Ground Movements	Bailey et al., 2016
An innovative Verification Process speeds construction of Crossrail's Moorgate shaft	Farooq et al., 2016
The Importance of Construction Mock-ups and Trials	Moxton and Atherton, 2014
Data management, analysis and visualisation on Crossrail Drive X Western Tunnels	Rendel and Thurlow, 2014
Design of Crossrail's Precast Tunnel Linings for Fire	Ryburn et al., 2014
Design of Sprayed Concrete Linings in Soft Ground: A Crossrail Perspective	Su and Thomas, 2014
Stepney Green Cavern: Design Concepts and Performance of SCL Lining	Uhrin et al., 2014
Design of Crossrail Farringdon Station: from an Engineer's Perspective	Kumar and Sharples, 2016
Effective MEP Design Techniques and Strategy for Large Subsurface Metro Projects	King, 2018
Application of stair pressurisation in Crossrail stations, shafts, and portals	Fung, 2017
Use of Non-Metallic Materials in Ventilation Airways- a review of aerodynamic performance and material requirements	Fung et al., 2017
How Will Rooms Operate on Loss of Mains Power?	David, 2017
Mechanical Design Principles	Matthew et al., 2017
Systems architecture models: Crossrail design and delivery from the client's perspective	Li and Georgiou, 2017
Examples of Specialist Grouting on Contract C310 Thames Tunnels	Smith, 2015
Compensation Grouting	Frances and Hayman-Joyce, 2015
Application of the Observational Method in Crossrail Projects	Chen et al., 2015
Maintenance and Cost Analysis on Major Subsurface Rail Projects - Crossrail	Consterdine and King, 2018
An Investigation of Surface Settlement and Volume Loss Associated with SCL Tunnelling at Stepney Green	Ssenyonga, 2018
Protection against fire for the UK Crossrail tunnel structures	King, 2018
Performance benefits associated with elevated temperature set points and increased error deadbands for room temperature controllers	Morgan and Fung, 2018

Table 2.4: Crossrail case studies on major projects using BIM

Title	References
Better Information Management to Optimise Whole Life Business Decisions	Dentten, 2018
Implementation of Energy Saving Initiatives in Design Review	Fung, 2018
Control of Railway Induced Groundborne Noise and Vibration from the UK's Crossrail Project	Cobbing et al., 2018
Client Advice and Technical Services to Chief Engineer's Group	Stowell, 2018
Earthing and Bonding	Anderson et al., 2018
Crossrail Project: Application of BIM (Building Information Modelling) and Lessons Learned	Taylor, 2018
Managing Reliability Growth in Practice	Consterdine et al., 2021
Final Design Overview - A Crossrail Design Assurance Process	Morgans, 2021
A review of reactive electrical power and power factor correction (PFC) in the Crossrail Low Voltage (LV) systems	Abu-Qulbain and Hill, 2021

Source: Crossrail Ltd. (2022)

These case studies provide a good legacy that the technology can be adopted in different projects and contribute towards the realization of positive results and quality project outcomes. This is because BIM facilitates in the attainment of faster and effective process, automated assembly, production of better quality, support of good design, control of the environmental data and whole-life costs, and realization of good customer service (Tookey, 2012). Azhar (2011) explored the case study of Holder Construction Company, in Atlanta, Georgia, which provided evidence that BIM facilitates in cost savings, increasing return on investment, better communications, and attainment of collaboration among project stakeholders.

The adoption of BIM is growing as owners are increasingly realizing its benefits. Studies have identified several notable benefits of BIM in building construction including better design, more effective processes, quick turnarounds, controlled life-cycle costs, improved quality and customer services, and more predictable environmental (and sustainability) performance (Azhar, 2011). Also, BIM has been associated with potential benefits in infrastructure projects. Nonetheless, the implementation of BIM in infrastructure projects is still a challenge. The following section explains the tangible advantages of deploying BIM in infrastructure projects.

Having reviewed the learning legacies of major projects, it is evident that BIM execution has a huge potential in major construction project delivery. The legacies depicted by Crossrail Ltd. (2022) are true evidence of the capability of BIM to support the construction works and realization of positive results. These facilitate in ascertaining that positive outcomes are obtained within the construction sector.

2.5 Technical Solutions for Performance Enhancement in Major Construction Project Delivery

2.5.1 Overview

BIM is utilized in major construction project delivery to provide technical solutions in cost management, safety and quality management, field management, and technical management, and recent analyses have ascertained that there have been significant improvements in BIM projects' communication efficiency (Chen, 2020; Pan et al., 2021). Furthermore, BIM is utilized in detecting clash analysis during the phase of design and optimizing the project cost, quality, and schedule (Waddoups, 2014). Patent analysis, SNA, and text clustering of patents are utilized as TM to obtain data distribution information on technical solutions of BIM, assisting in the identification of the BIM application hotspots and trend forecasting. Pan et al. (2021) demonstrated that BIM technology is undergoing massive development, which allows it to be applied in different sectors and regions of the construction industry. The approach ensures that it is possible to design a building maintenance system that is efficient (Meyer, 2016).

According to Choo et al. (2019), new technologies that include BIM tools often pose a challenge on matters of intellectual property, ownership, cultural differences, and misunderstandings. These issues have the potential of leading to other problems linked to aspects of data generation that include errors, data discrepancies, data loss, and liability for incomplete or incorrect data (Lam, 2004). However, conducting a risk-benefit analysis of BIM reveals that the technology has positive benefits in the construction industry (Ma, 2019). The benefits are realized when nations undertake the step of having policy changes related to copyright problems, standards and guidelines, and cost issues that act as core obstacles for the practical adoption and implementation of BIM (Li, 2019).

As opined by Gopalakrishnan et al. (2017), BIM is inherently data-centric, and it has long been acknowledged that the AEC industry has to comprehend aspects of Big Data processing, storage, and visualization challenges (Stenlund, 2005). This is because of the variety and huge volume of the information that is managed within BIM, alongside other sources integrated into mainstream BIM paradigms such as energy management and geographic information systems (GISs), which are integrated to BIM (Khan and Ahmed, 2012). Often, a cloud-based framework that offers web-based service for storing, processing, and viewing massive dynamic BIMs is adopted to overcome the computation burden of managing multiple BIM projects (Stakeholder Democracy Network, 2014). The cloud-based application of BIM provides technical solutions, which ascertain that there is better performance in the delivery of major projects. The approach also ascertains that it is possible to assess and evaluate the project lifecycle efficiently (Gopalakrishnan et al., 2017).

2.5.2 The Need for Advanced Solutions in Major Construction Projects and Delivery

Major construction projects, as explained previously, are projects that are associated with extensive complexities, incurring high costs and integral technical requirements. The endemic cost overruns and delays that are considered normative in major construction projects result from their inherent complexities and uncertainties, which in turn entail long deadlines, and the size of such projects affects the productivity of various project teams (Bendale, 2018). The successful management of major construction projects necessitates the exploration of projects factors, including (but not limited to) project size, value, realizable goals, complexity, type, and risks. Given that project factors can potentially impede project success, understanding this factor and formulating feasible solutions to looming risks is cardinal in warranting success in the construction world.

Managers of major construction projects ought to consider the project type; technology, pace, novelty, and complexity; the competence of the organization in project management, market value, the level of trust, and the stage that the project is in in reference to the project lifecycle as a mean to propel the project towards its goals. Bendale (2018) affirmed that the study of project factors before the commencement of the project is cardinal to ensuring that desire quality levels and project success are gained. Projects in the construction industry are substantively different to major projects in other industries. The unique factors and characteristics of projects in the construction industry introduce complexities that elevate the predisposition of projects to the occurrence of risks and project failure.

The design stage of every major construction project is an integral stage where feasible solutions to eminent risks can be determined, minimizing the costs of rectification (e.g., by amending designed before physical construction is underway) and the likelihood of project failure. Nonetheless, the design stage is inherently iterative, and it is dependent on multidisciplinary coordination and the team's competence (Badran et al., 2020). Inadequate design interphase management often results in the proliferation of project implementation risks, project duration overruns, poor-quality of the deliverable, cost overruns, and the failure of the project to meet set goals. Although a vast assortment of design information management tools has been developed in the recent past, most of the available tools are not used in the construction industry, except the BIM. The hugely documented benefits of BIM to the construction industries necessitate an exploration of how the BIM can be used to effectively manage project risks.

To effectively decipher the efficiency of BIM in the management of design risks, evaluating existing tools such as the BIM-based risk management framework designed by Badran et al. (2020) can be instructive. Their solution was designed to propagate effective management of the design stage-related risks for a high-rise building belonging to design-bid-build (DBB) in the UAE. The results deduced from using the BIM-based risk management framework denote that the framework significantly increased the chances of project success. Nonetheless, the adoption of the framework was dotted by several challenges relating to the lack of adaptable standards, interoperability, and security. According to Mahamadu et al. (2013), an increased perception of operational risk at the expense of actual challenges that could be encountered during the project implementation stage has cause a degree of reluctance in the usage of BIM as a platform for informational exchange. Moreover, Chen (2019) proposed that "*when considering major construction projects, the main obstacles can be classified into three groups: workforce/people, production, and processes*."

In addition, technical challenges encountered in the implementation of mega construction projects can be divided into four categories. The first category is sound or rational professional competence pertaining to workforce and competitive leadership which can be attained through effective usage of the Construction Management Body of Knowledge (CMBOK) (Chen, 2019). The second challenge is reasonable budget alongside effective process control pertaining to the allocation and usage of resources. The third challenge is reliable schedule, and this entail efficiency in the usage and control of time. The last challenge is substantial sustainability (Chen, 2019).

Chen et al. (2020) assert that a practice-oriented requirement for researchers looking into the management or mega construction projects and professionals involved in the planning and implementation of the project can be instrumental in tackling major issues. The technical solutions that are usually informed by effective synthesis of data and professional knowledge are crucial in allowing stakeholders to achieve strategic value from projects (Bishop, 2012; Ewejea et al., 2012; National Audit Office, 2013). The bottom line is that reliance of information that can be generated through the effective use of technological solutions is crucial in steering the design and implementation of major construction projects towards project success.

2.5.3 The Use of Digital Solutions in the Construction Industry

Technology has become an integral part of today's digital economy, and is increasingly ubiquitous in almost all spheres of life; therefore, the construction industry is not immune to the looming technological revolution. The adoption of technology in the construction industry had been necessitated by the prevalence of age-old questions, problems, and risks that have cause the failure of numerous mega construction projects. Myriad technologies are availing potentially effective solutions to long-standing issues in the time overruns, cost overruns, issues in the coordination of multidisciplinary teams involved in the planning and implementation of the projects, and the ability to meet project specification and deliverable set by the client. Nonetheless, Ibrahim (2013) asserts that further studies about the implementation of these innovative solutions must be conducted in an attempt to separate the myth surrounding the effectiveness of these technological solutions from the reality.

Although technology is increasingly being adopted in the construction industry, a wholly integrated approach has not been realized in most construction projects. According to Ibrahim (2012), innovation has been adopted to propagate partial integration between the workforce and technology. As a result, the full benefits of technological solution as they apply to the construction industry have not been realized. Moreover, digital solutions that propagate sustainability in the management of mega construction projects have received somewhat less attention despite the fact that governments across the world have increasingly become interested on how sustainability can be achieve through the use of innovative solutions. Furthermore, innovative solutions are used differently from project to project, and this signifies that a uniform approach is yet to be determined. The diversity in the usage of digital solutions in the construction industry implies that there are no best-practices pertaining to the use of these digital solutions. As a consequence, some construction projects are bound to fail due to wrong integration of digital solutions.

The difference in the implementation of digital solution to the mitigation of life-long problems encountered in mega construction projects means that even the most used model, BMI, means different things to different PMs. Even though there is an industry-level adoption of BIM, the perpetual question of interoperability still persists, necessitating the advancement of a broader definition of BIM to encompass the technical elements. Ibrahim (2013) also identified affordance, technologies' material constraints, information-risk, measurement of value, and leadership as some of the persistent challenges faced by PMs. In an attempt to elucidate how technology or innovation can be used to address some of the aforementioned challenges, he called for further studies should focus on optimal implementation of BIM, especially in the context of mega construction project management, the challenge of integrating different technological tools across the entirety of the project lifecycle, and how technology can be used to drive sustainability in the management of major construction projects.

Although BIM has been hugely adopted in various stages of the lifecycle of most major construction projects, the wide range of required technical expertise required to successfully use the technology and the complexities of the technology and the major construction projects still poses a challenge to industry-level adoption of BIM. Retzlaff

(2018) indicated that by employing optimal strategies in the integration of the technology, the goals of BIM can still be realized, and he demonstrated that cooperation among parties involved in the planning and implementation of the project when using data derived from a single BIM model requires proper delineation of design and planning responsibilities.

The review on technical solutions for performance enhancement indicates that BIM has a huge potential in supporting project performance and productivity. As such, understanding of the BIM application hotspots and trend forecasting facilitates in supporting the realization of positive results in the innovations for the construction industry. Hence, BIM has a huge potential in creating benefits that can be enjoyed by stakeholders within the construction industry.

2.6 Industry Guides and Standards for BIM Executions

The execution of the BIM adheres to set protocols that help in prescribing how BIM needs to be utilized by service providers, which are tied to the construction contracts to ascertain that value is delivered. Government agencies have embarked on the creation of the regulatory framework that is vital in supporting and promoting the adoption of BIM through national standards. The majority of existing standards are descriptive (taxonomy-based) guides, while others are prescriptive (protocols in taxonomy) (Sacks et al., 2016). The standards or guidelines for BIM execution focus on construction process, design phase, collaborative practice, technology for use, and the information that has to be provided for the process (Bannister, 2004). However, the guide, protocol, and mandate for BIM differs from one place to another, based on standardization approach, specificity and requirements for technology, scope, level of information specifications, and definition of construction process and design (Queensland Government, 2021).

Industry Foundation Classes (IFC) is one of the primary standards supporting BIM execution. Developed by buildingSMART, IFC is an open, international standard that enables data exchange and interoperability among various software applications within the AEC industry. It serves as a foundational structure for BIM, providing a common data schema to represent building and construction industry data across different BIM applications. IFC supports collaborative workflows by enabling seamless data exchange between disciplines, ensuring that project data remains consistent, accessible, and usable by all parties throughout the lifecycle of the project (buildingSMART, 2020). Its widespread adoption has proven essential in achieving interoperability in BIM projects, facilitating communication and data integration that reduce errors and improve project efficiency.

Uniclass, another vital standard, provides a unified classification system for organising information across all aspects of the construction industry. Developed in the UK and maintained by the National Building Specification (NBS), Uniclass structures project data by categorising elements of buildings, infrastructure, and construction processes, making it easier to manage and retrieve information across project stages. Uniclass's flexibility allows for detailed classification, covering a wide range of aspects including entities, activities, systems, and products, which aids in organising BIM information and supporting clear communication within project teams. By integrating Uniclass, organisations can standardise data categorisation across projects, aligning with international standards and streamlining the flow of information, particularly in large and complex construction projects (NBS, 2021). Uniclass complements IFC by providing a consistent way to classify and structure BIM data, further enhancing collaboration and efficiency.

There are 81 BIM guidelines, which are included in BuildingSMART's BIM Guides Project map (Karlson and Ronndahl, 2018). The project map acts as a good resource for the review, modification, and creation of BIM guides. Several BIM documents have guidelines and standards that have to be followed in the process of implementing BIM (Martin, 2017). Different official bodies have BIM standards, which act as the guidelines and procedures for the management of the project lifecycle (BibLus, 2022).

The core international BIM standard that plays the role of regulating BIM execution in AEC industry is ISO 19650 (BibLus, 2022). The standard has five parts that include: general principles and concepts, phase of delivery for real estate assets, information management and development processes, exchange of information, and requirements for information security (Chen et al., 2015). Such standards and guidelines are necessary in execution of BIM to ensure that the entire system is organized and has a good order (Cao et al., 2019). Unfortunately, not all countries have such provisions, since the regulation of BIM has not yet occurred at the national level (Erdogan et al., 2010). Countries that have adopted BIM and have these standards have been able to record positive performance in their projects, including quality outputs.

The review of industry guides and standards for BIM execution indicates that the technology holds significant potential for improving the performance of major project delivery. These standards are instrumental in enhancing overall project quality, reducing timelines, increasing efficiency, and optimising processes. However, while many existing standards focus on structured data formats, the current research addresses the handling of unstructured textual data that traditional BIM standards often overlook. By integrating this unstructured textual data within the BIM framework, this research aims to unlock additional insights and further enhance the performance of major project delivery, highlighting an often-underutilised dimension of BIM's potential.

2.7 BIM Execution Potential in Major Construction Project Delivery

BIM has been introduced in the construction sector as a core mechanism that helps in addressing issues of inefficiencies, which are related to methods of project generation and delivery. The execution of BIM increases the collaboration level in the execution of the project, which in turn has positive influence on building projects completion time and quality (Gercek et al., 2017). BIM plays the role of overcoming traditional limitations of human actions by introducing aspects of improved quality and timely delivery of projects (Carvalho et al., 2019). The situation occurs since it involves the combination of technology and processes to improve effectiveness and efficiency level of project delivery from stage of inception to maintenance and operation (Latiffi et al., 2016).

BIM has potential in construction projects on basis of quality, cost, and time. The technology also generates design clashes analysis that is vital to the design team since it assists the team in solving the clash issues during early stages of construction works (Wiedemann, 2013). These activities ensure that there is maximum avoidance of construction cost overrun and project delay issues (Latiffi et al., 2016). Clients have the chance to visualize and understand the project through BIM, which improves their level of satisfaction with the project. Furthermore, project contractors achieve effective site coordination and collaboration when they use BIM (Andriamamonjy et al., 2019). The strategy also ascertains that effective safety plans are adopted for the construction sites to avoid any form of mishaps.

According to Zhen-Zhong et al. (2022), BIM technology concept focuses on the application of digital modelling and associated technologies (DMAT) as a means of adopting technique and collecting data. The readable data is in standard format such that it can be shared among participants in various phases of the project lifecycle (Huang et al., 2020). BIM also integrates domain knowledge and expert methodologies for the intelligent and automated applications (Zhen-Zhong et al., 2022). These support BIM-based knowledge that improves efficiency and generality of knowledge management.

Hodorog et al. (2021) suggest that leveraging text-mining and clustering techniques provides a valuable opportunity for gaining insights into the implications of BIM within the supply chain. Additionally, these methods offer the potential to update and enhance existing skills and competencies in this domain. The approach ensures that AEC industry has to addresses challenges of project execution, which are tied to linear workflow, risk avoidance, and adversarial relationships. Thus, BIM plays a vital role in strategic agenda to comprehend how the AEC industry functions, including having BIM training protocols aligned to the ever-changing evolution and demands of the industry (Foxe, 2010).

The review of the potentials of BIM execution in major construction project delivery indicates that the technology is vital in the construction sector. Furthermore, it is evident that BIM integrates with TM and other technological developments to generate positive results in the sector. Hence, it is paramount to focus on ascertaining that the technological tool is adopted in society to achieve positive outcomes in construction works.

2.8 Applied AI for Major Project Delivery

Smith and Wong (2022) described AI as the development and theory of computer systems to enable the systems to perform intelligently including such tasks as visual perception, decision-making based on certain inputs, speech recognition, and similar human tasks. The AI applications allow for decision-making in project managements (Arayici et al., 2012). ML is one of the AI applications that involve the development of computer programs where past data is learned to make predictions (Lang and Baehr, 2012). Several algorithms are involved in ML such as logistic regression, artificial neural network, multivariate-linear regression, K-means, random forest, Bayesian inference, decision tree, and support vector machine (Pärn et al., 2017). These facilitate in ensuring that the project has the opportunity of achieving success when AI project management is adopted.

Fuzzy logic depicts a technique that is adopted in measuring the level of correctness for data that is uncertain. The technique is adopted in real-world systems, where it helps in tackling complex and ill-defined problems (Smith and Wong, 2022). NLP is a tool for AI that focuses on creation of computational models, which are equivalent to linguistic capability for the human beings. AI also has the evolutionary algorithms, which is related to numerical optimization, decision support, and biology applied to different applications in engineering (Ruiz et al., 2021). The AI technique and tools ensure that BIM has the capacity to offer decision support system that is needed to ascertain success in the delivery of the project (Connolly-Barker et al., 2020).

2.8.1 AI as Applied in the Construction Sector

McCarthy and Hayes (1968) note that the term AI was introduced at the Dartmouth conference. This event was organized by researchers who held a specific fascination with intelligent machines. To tackle fundamental inquiries surrounding AI, including topics like neural networks, self-improvement, automated computation, computer language utilization, and abstraction, major challenges within the AI field were taken into consideration The designation "AI" (Russell and Norvig, 2010) is specifically attributed to machines possessing a certain level of intelligence within a limited scope. This pertains to devices that replicate the processes of human decision-making and reasoning in order to effectively address problems. AI’s impact cuts through many fields and technologies, including game play. In the construction industry, the automation that is associated with AI could potentially optimize the delivery of major construction projects.

2.8.2 AI for Improving Construction Projects Delivery

Though there are different views on AI's potential in construction, it is unavoidable that AI will introduce numerous changes to the positions of various workers working in the industry. AI has the potential to significantly reduce the workload of architects by automating redundant tasks and facilitating the seamless conveyance of critical project information, enabling more efficient project management and decision-making (Shamim, 2024). This capability can streamline workflows, allowing architects to focus on design and innovation while AI manages data-heavy tasks. Nevertheless, Blanco et al. (2020) argued that AI should never be perceived as a risk that will remove the need for humans from the human resource facet of construction companies. Alternatively, it can be viewed as an integral too that can improve the skill sets of professionals while maintaining a smooth workflow and building successful buildings.

2.8.3 The Current State of AI in Construction and Engineering

Even though a broad array of literature has explored how the adoption of AI in the construction industry can be integral in propagating successful major construction project delivery, the trend is not yet at an industrial level. Nonetheless, start-ups in the construction industry is more likely to adopt AI than already established firms, for whom existing and time-honoured ways of doing things have proven effectiveness (while AI and even BIM adoption can be perceived as hazardous and costly adventures) (Blanco et al., 2020).

Due to inherent inhibitors of change adoption (i.e., legacy systems, inertia, and resistance), and the inherent barriers to new (i.e., generally more innovative) entrants in the AEC industry, the adoption of AI solutions alongside other ML technologies is relatively low in the sector worldwide. Blanco et al. (2020) analysed a McKinsey Research report which compared building materials and construction projects to 12 other industries, and elucidated that in terms of current AI adoption, ten of those industries are better placed to succeed because they are miles ahead as far as AI adoption is concerned, but all 12 were expected to increase spending on AI in the coming years at a faster rate.

Notably, Blanco et al. (2020) explain that AI algorithms learn from past data, so having a large amount of data is crucial for effective AI use. Companies with access to extensive project data are better equipped to train these algorithms. As a result, in the near term, larger corporations with more data resources are likely to see the greatest benefits from using AI in construction (Blanco et al., 2020). An external third party may join and exploit firms' knowledge to train its models, a scenario that would likely lead to enhancement for individual companies across the industry as a whole, but which would limit competitive advantage derived from investment in such innovations, but this might be viewed as an exaggerated corporate concern given the enormous restrictions on data sharing and conventional data ownership regulations pertaining to AEC (Blanco et al., 2020).

2.8.4 Type of Applied AI Systems

Numerous computer models have found application within the construction industry for diverse purposes. For instance, they have been employed in tasks like value management (Cheng et al., 2009), forecasting occupational safety risks (Tsoukalas and Fragiadakis, 2016), estimating task durations (Hong et al., 2011), performing built analysis (Yu and Skibniewski, 1999), and predicting instances of bankruptcy (Alaka et al., 2016; Jackson and Wood, 2013).

Based on the analysis of existing literature, AI-based intelligent model creation approaches widely used in the construction sector can be classified into four groups: (a) ML techniques (b) evolutionary techniques, and (c) knowledge-based techniques, and (d) hybrid techniques. Figure 2.3 shows the comparison of the four AI groups, and Figure 2.4 demonstrates AI techniques in construction studies.

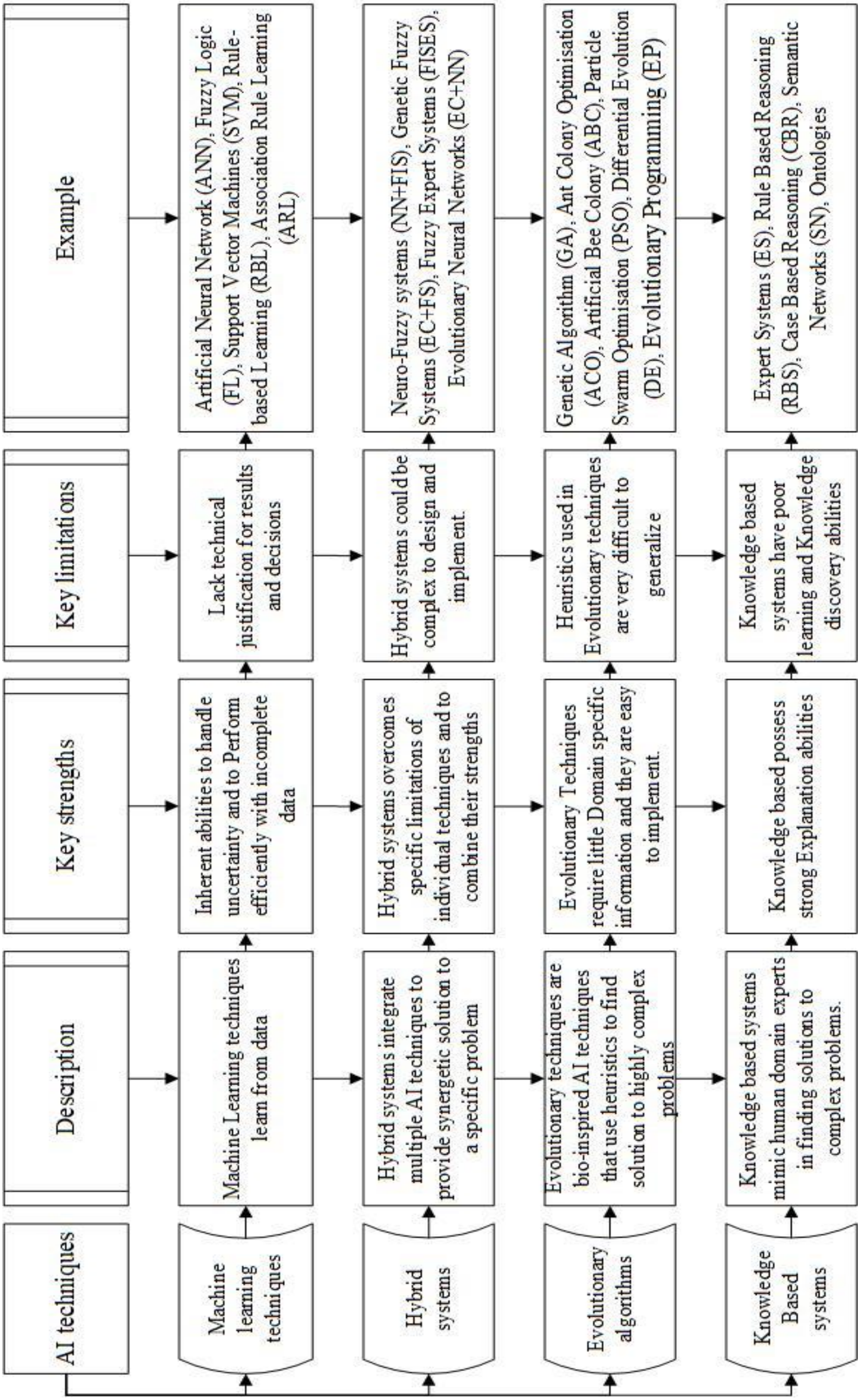


Figure 2.3: AI techniques in construction studies (Chapman et al., 2000)

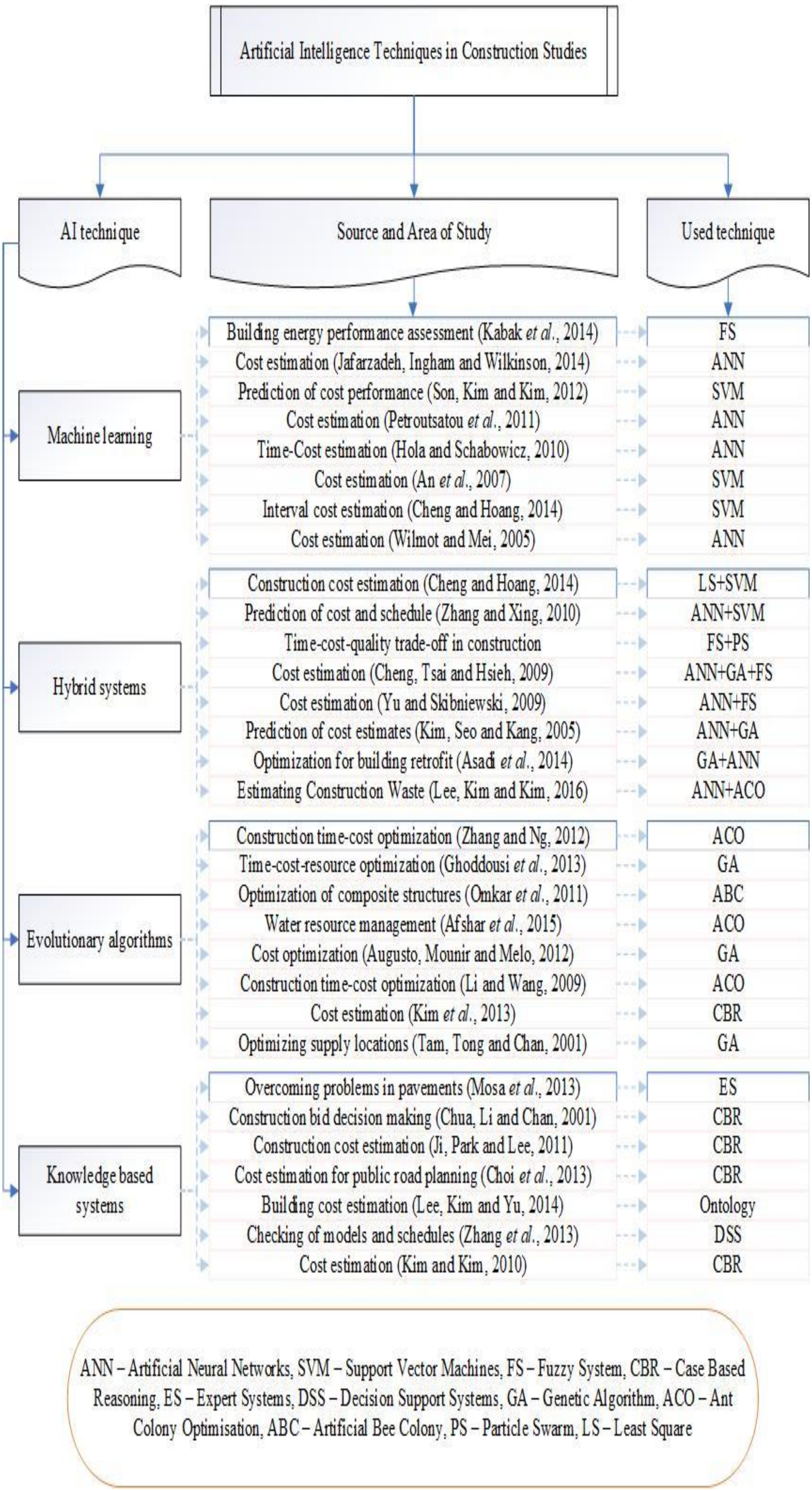


Figure 2.4: AI techniques in construction studies

2.8.5 Modelling of AI-Powered BIM

In recent years, AI has increasingly influenced a wide range of industries, revolutionising processes and transforming market dynamics. AI applications are now prevalent in sectors such as healthcare, finance, retail, and manufacturing, where they improve efficiency, enhance decision-making, and create new opportunities for innovation (Davenport and Ronanki, 2018; McKinsey and Company, 2020). In the context of the construction industry, BIM is garnering almost similar traction. BIM operates on a fundamentally integrated network that facilitates the exchange of knowledge, raw data, and processed information, enhancing project visibility and collaboration across stakeholders. This interconnected structure allows for seamless information flow, supporting decision-making throughout the project lifecycle (Eastman et al., 2011; Succar, 2009), even in the case of major construction projects. BIM has enabled engineers, architects, designers, and contractors to work closer together in an increasingly collaborative manner and work for the broader picture. AI technology is pervasive across sectors, and it may also contribute to optimal project delivery in the construction sector (Blanco et al., 2020).

BIM software, as used in both major construction projects and conventional building projects, creates a large volume of data that is hard to handle manually. Consulting companies have set up whole departments dedicated to the processing, storage, and organizing of information generated through the use of BIM. Data obtained from multiple sources, such as models, simulations, and sensor-like physical components, can help transform the design process and optimize the entire process of project management (Blanco et al., 2020).

2.8.6 Methods of Implementing Machine Learning (ML)

According to Russell and Norvig (2010), intelligent approaches that can learn from data are ML techniques. Due to their inherent capabilities to deal with uncertainties and to achieve objectives with incomplete data, ML techniques have become very common, particular in computer science studies. ML approaches operate by using gained perspectives from related cases to judge new cases. Nevertheless, a significant drawback of machine learning methods is their inherent deficiency in providing a clear technical rationale for the results and decisions they produce, notwithstanding blackbox structures called ML techniques (Russell and Norvig, 2010). Despite this drawback, in the building industry, ML techniques have been commonly used.

Artificial Neural Network (ANN), Support Vector Machines (SVM), Fuzzy Logic (FL), and rule-based ML are the most popular among the various ML approaches accessible to players in the construction industry. However, ANN, SVM, and FL tend to be the most frequently utilized in the construction industry (Irani and Kamal, 2014). ANN is an ML technique that uses a series of artificial neurons to mimic the cognitive capacity of the human brain. Evidence from the available body of literature shows that ANN generally outperforms SVM in construction-related tasks (Kim et al., 2013; Zade and Noori, 2008). Therefore, considering what works best can be used to optimize the process of major construction project management.

Deep learning, because of its ability to extrapolate advanced functionality from a limited collection of functions, is becoming popular. Deep learning is revolutionizing the realm of AI, cantered around the concept of neural networks. It empowers the creation of computational models featuring numerous layers of processing, designed to grasp the formatting and representation of data through multiple tiers of abstraction. Within the scope of Big Data, this approach aims to enable the recognition of intricate structures. This involves adjusting the internal parameters to compute the representation within a given layer based on the representation in the preceding layer.

Various deep learning architectures exist, including evolutionary deep neural networks, deep neural networks, deep belief networks, deep Boltzmann machines, long short-term memory models, and recurrent neural networks. Another noteworthy machine learning technique that has gained prominence in recent times is Association Rule Learning (ARL). Since its introduction by Agrawal et al. (1996), ARL has found applications in diverse domains, with a particularly notable application in the realm of market basket analysis. The fundamental idea underpinning association rule learning is to unveil common traits among elements within a database. Its objective is to identify both resemblances and disparities within collections of objects stored in the database.

After reviewing the applied AI solutions in major construction project delivery, it is clear that integrating TM with BIM have the potential to enhance project performance in major construction projects. By leveraging TM to process unstructured data, BIM can not only improve decision-making but also optimise workflows and streamline the management and scheduling of project activities. This integration supports data-driven decisions that promote efficiency, address the complexities of large-scale projects, and contribute to achieving higher levels of sustainability.

2.9 Theoretical Advances in Data, Information and Knowledge (DIK) Applications

2.9.1 Theoretical Basis

The relationship between data and knowledge has been demonstrated using different information systems theoretical perspectives. Other disciplines have also contributed to the understanding of the relationship, and different definitions of key information system concepts are thus common. In explaining the theories and definitions of these key concepts, Li and Kettinger (2022) provided examples for a production planning context, as tabulated in Table 2.5.

Table 2.5: Testing the relationship between information and knowledge

Concepts	**Working definitions**	**Production planning examples**
Data	Measurements of entities or occurrences.	Examples include customer demand, material cost, labour cost, product types, etc.
Knowledge	A dynamic blend of structured experiences, values, and informed judgment that offers a framework for incorporating new experiences (Davenport and Prusak, 1998).	Rules for production planning, such as balancing production with demand, minimizing overhead, and managing inventory.
Information	The interpretation of data to choose a different course of action (Kettinger and Li, 2010).	Vital details like product inventory levels and production overtime. Adjustments to the production plan in case of shortages, excessive inventory, or overtime.
Domain-knowledge	Also known as field-specific knowledge, it's the information a person gains from a particular problem domain (Carlile, 2004; Hicks et al., 2002).	General comprehension of production planning, encompassing concepts, rules, and procedures.
Task-knowledge	Knowledge specifically used in a decision task (Barrick and Spilker, 2003; Deng et al., 2008; Thibodeau, 2003).	Specific insights pertinent to production planning tasks, such as factors influencing inventory, material and labour costs, and the implications (e.g., profit/loss) of overtime.
Design Stage	The phase of decision-making where data, knowledge, and information are gathered for problem structuring and creating alternatives (Browne and Pitts, 2004; Frisk et al., 2014).	Input elements such as customer demand, material and labour costs, used to formulate various plans using existing knowledge.
Choice Stage	The stage in decision-making when information is derived from data and knowledge and applied to make selections or refine decisions (Browne and Pitts, 2004; Frisk et al., 2014).	Essential information such as production levels, inventory, and overtime deployed for decision-making and enhancing production plans.

Source: Li and Kettinger (2022)

While commenting on the relationship between data, information, and knowledge, Kettinger and Li (2010) observed that three generic views had been documented, of which the Data-Information-Knowledge (DIK) hierarchy founded by Ackoff (1989) was presented as the most popular. This hierarchy claims that data facilitates both information and knowledge. Secondly, the inversed KID view was argued for by

Tuomi (1999), postulating that knowledge was a result of both data and information "i.e., information is articulated knowledge, while data are structured information". The interactive perspective asserts that information emerges from a combination of both knowledge and data (van der Spek and Spijkervet, 1997; Langefors, 1980). Different arguments have developed concerning these variant views.

While the DIK hierarchy has been considered popular in information system research, several limitations have been brought forward. Kettinger and Li (2010) observed a case where it was possible to get different information from a single dataset. Other observers have indicated the difficulty in distinguishing between data and information (Shin et al., 2001; Tuomi, 1999), and separating information from knowledge. Following the noted limitations, various perspectives have been suggested, and among these is the concept of a circular connection between data, information, and knowledge (Knox, 2007). Moreover, there is even a proposition for a fundamental shift in paradigm, moving away from mere information processing toward actively generating knowledge. Such an approach, according to Nonaka and Toyama (2005), prevents overshadowing the key role of information system.

The perspective of interaction, as demonstrated by the knowledge-based approach to information proposed by Kettinger and Li (2010), could potentially offer a solution to this challenge. This standpoint implies that information emerges from the interplay of knowledge and data, with knowledge dictating the type of information that can be derived from data. Knowledge encompasses a dynamic fusion of structured experiences, values, and contextual and expert insights, serving as a foundation for assessing and assimilating novel experiences (Davenport and Prusak, 1998). It outlines the connections between entities or ideas (Brydon and Gemino, 2008), including production guidelines, decision trees, and even sophisticated algorithms.

In the case of individuals, knowledge finds expression through their cognitive capacities or mental frameworks, whereas in decision support tools, knowledge is ingrained within computerized decision models (Baker, 2013; Watson, 2018). Data, on the other hand, serve as the fundamental elements for generating information, signifying the measurements of entities or occurrences that convey data for processing.

It's noteworthy that identical data can lead to varying forms of information for distinct individuals or systems, contingent on the knowledge or decision models utilized.

2.9.2 Data Science Skill Dynamics

In order to allow data scientists to travel between sectors, collaborative, sustainable methods are required (Blake, 2019). In order to accomplish the vision of the UK as a leading data science research nation, Blake (2019) identified four primary action areas with recommendations for addressing priority requirements across the data science talent pipeline, from school to advanced professionals. As a result, data has the potential to considerably boost the UK's economic competitiveness and productivity, both through new data-driven business models and existing organizations adopting data-driven procedures (Dowden, 2022).

However, Dowden (2022) further emphasized that existing evidence supports broad economic benefits from better data utilization, including a relationship between efficiency, productivity, and data-driven corporate practices. Proficiency in data skills holds advantages for all individuals; enterprises that effectively harness data will gain a competitive edge in the contemporary digital economy. Similarly, data-literate people are more likely to profit from and contribute to increasingly data-rich surroundings, while data-driven businesses can significantly boost their and the economy's productivity.

Data skills are in high demand across the economy. Following the Department for Digital, Culture, Media and Sport (DCMS)-commissioned examination of 9.4 million internet job advertisements, the Royal Society expected that data analysis abilities would be the fastest-increasing digital skill cluster into the mid-2020s (Blake, 2019). This reflects the exponential demand for advanced data science and ML applications across all industries, from cyber to construction. The emergence of AI and cyber specialisms increases the need for more basic data skills to feed the advanced skill pipeline and offer businesses the skills they need to operate with data. Notably, these limited skills were critical in the coronavirus response. Their contribution to UK RandD is important, and is expanding rapidly.

2.9.3 DIK and BIM

Within the framework of the knowledge-based perspective, information is construed as a representation of an individual's intention and willingness to act, rather than just the meaning of data (MacKay, 1969). This view considers information as the result of a thought process that takes into account prior knowledge and current data and is seen as the direct input for decision making. This view provides a new way of understanding information and its role in decision making. An experiment was undertaken by Bond et al. (2007) in which individuals made purchasing decisions by considering product attributes. The experiment found that the same attributes can generate different information for individuals depending on their prior knowledge. This highlights the influence of prior knowledge and mental models on the interpretation of information and subsequent decision making.

The example given shows that the interpretation of information can vary greatly between individuals, even when presented with the same data. For instance, one individual might perceive the price of a product as expensive, whereas another person could view the same price as affordable. Similarly, one person might regard a product feature as significant, while another individual might perceive it as insignificant. This difference in interpretation is due to the influence of prior knowledge and mental models on the individual's understanding of the information. The knowledge-based perspective on information distinguishes it from both data and knowledge. It highlights the critical role of utilizing relevant knowledge to accurately interpret data, thereby generating appropriate information crucial for informed decision-making (Bond et al., 2007).

This view acknowledges that information is not just the meaning of data, but rather a representation of an individual's intention and willingness to take action, shaped by prior knowledge and mental models. By considering the role of prior knowledge and mental models in shaping information, this perspective provides a more nuanced understanding of the decision-making process (Pauleen and Wang, 2017), and highlights the difficulties in defining the boundaries between information, data, and knowledge. There is often a lack of clarity about the distinction between these concepts, and the relationship between them can be complex and difficult to define.

For example, some might define knowledge as distilled information, while others might see it as separate from information.

Similarly, the boundary between information and data can also be unclear, with some defining information as structured data and others seeing it as something distinct from data. These controversies surrounding the definitions of these key concepts demonstrate the importance of a clear understanding of the distinction between information, data, and knowledge (Tuomi, 1999; Shin et al., 2001). According to Li et al. (2022), BIM has been intensively adopted as a means of improving discontinuity and fragment of the supply chain management (SCM) construction. The kernel of BIM is information, which ascertains that construction stakeholders in SCM are informed of all operations through collaboration with one another (Shi et al., 2016). SCM has cross-relation of different data processes, which introduces the semantic information to the DIK paradigm.

DIK is used as a framework for the elevating and combining multimodal data to the models of knowledge and information (Miller and Stasiuk, 2020). The approach ensures that it is possible to map the evidence from the SCM practice, the system, and then the users. Furthermore, DIK depicts the ability of BIM to embrace information technology in a hierarchical manner, which results in gaining positive outcomes in projects. In practical terms, DIK conceptualizes BIM with regard to management requirements for ample and operational information availability (Dammann, 2019).

Data identifies symbols as collected, simulated, or retrieved from text-mining, measurements, survey results, sound recordings, or images. The data could be depicted or tabulated in graphs or displayed in form of figures. The presentation of data is in qualitative or quantitative values of variables. Information refers to the data that is contextualized (Jifa, 2013). Thus, information is regarded as the aggregation collection, presentation, and analysis of data that offers an understanding. Knowledge is testable, consistently successful, and predictive belief. The theory of DIK is presented as a wall, wherein each block is a step to a higher level (Dorji and Kirikova, 2012). These are true depictions of how BIM should function in the AEC industry.

2.9.4 Summary

The review of the theory of DIK indicates that BIM has a huge potential in construction industry, and BIM has to focus on embracing information technology including AI so that it can achieve the desired success level in AEC industry. The DIK framework also illustrates that decision-making is undertaken with the objective of attaining optimal quality results in the delivery of major construction projects. In addition, the data science skill dynamics will assist in improvement of the overall project quality, reduction of time, increase in the level of efficiency, and optimization of the processes. Furthermore, it is evident that BIM integrates with TM and other technological developments to generate positive results in the sector (Al Hattab, 2021). Hence, it is paramount to focus on ascertaining that the technological tool is adopted in society to achieve positive outcomes in construction works. Data science skills dynamics also depicts that decision-making is executed with the aim of attaining the highest quality results in the execution of major construction project delivery.

2.10 The Theory of BIMTAPE Framework for BIM pervasive major project delivery

According to Frey et al. (1999), textual analysis depicts a researcher’s communication method for interpretation and description of the characteristics of visual or recorded message. Textual analysis ensures that the structure, content, and functions in the text message are described. BIMTAPE Framework focuses on text analysis, which ensures that the message that is provided in the text is described. Thus, all the words that are found in the text linked to BIM pervasive major project delivery are analysed to offer insights of their function and meaning. The approach will involve the use of a software tool, which will transform the formal process of development to project format that the tool can work with (Kuhrmann et al., 2014).

BIMTAPE Framework will also ensure that it is possible to achieve text-mining. The process will help in getting to understand that BIM has potential in major construction project delivery. For BIMTAPE Framework to be successfully implemented, it is necessary for the systematic literature review to give a brief knowledge on how TM was applied in construction projects. In order for BIMTAPE Framework to be fully implemented, it is empirical to see the level of adopting TM in AEC domain along with the level of adoption of TM methods in the construction sector.

2.10.1 Integrating Knowledge Graphs for Enhanced BIM Practices

The integration of Knowledge Graphs (KGs) in BIM practice provides a structured framework that improves the representation and handling of complex relationships within building information. Knowledge Graphs offer a rich, interconnected structure that captures semantic relationships between entities, allowing BIM data to be organised in a way that enhances both comprehension and usability (Pujara et al., 2013). Unlike traditional BIM data management systems, which often struggle with heterogeneous data sources and disjointed information silos, KGs enable the creation of an ontological structure, providing a unified schema that is especially effective in domains requiring high interdependence between diverse data types (Galkin et al., 2017). By mapping BIM concepts, standards, and terminologies within a knowledge graph, BIM practitioners can retrieve more contextually relevant information, making KGs a foundational tool for advanced data retrieval and analysis within the BIM domain.

2.10.1.1 Supporting Contextual Reasoning in BIM through Graph-Based Knowledge Representation

Knowledge Graphs are particularly beneficial in supporting contextual reasoning over BIM data by leveraging their capacity to structure and interlink various pieces of information. In BIM, where practitioners often need to understand complex and interdisciplinary relationships between building components, regulatory standards, and environmental data, KGs provide an advanced approach to linking data points that would otherwise remain isolated (Buehler, 2024). Through a process of 'graph reasoning', KGs can uncover implicit relationships across datasets, offering insights that can improve decision-making and problem-solving in BIM projects. For example, within a BIMKG, interdependent elements like structural components and materials can be connected to relevant regulations, enabling more effective risk assessments and compliance checks. The ability of KGs to facilitate reasoning and discovery over complex datasets highlights their value as a tool for innovation in BIM.

2.10.1.2 Enhancing Knowledge Accessibility and Discovery in BIM through Knowledge Graphs

The use of Knowledge Graphs in BIM enhances accessibility to diverse forms of building data, facilitating efficient querying and knowledge discovery. KGs allow for intuitive and accurate querying by connecting terms and standards that share semantic relationships, thereby supporting context-aware searches across extensive BIM documentation. For instance, using KGs in BIM could enable a user to trace standards associated with a particular material or building component, thereby speeding up the retrieval of relevant data and reducing redundancy in data searches (Buehler, 2024). Moreover, the ability to identify clusters of closely related BIM data points, such as similar construction methods, materials, or safety standards, enables BIM professionals to extract meaningful patterns that inform future designs and optimisations. The modular and scalable nature of KGs, which often exhibit high modularity and community structure, makes them particularly well-suited for managing the complex and highly interconnected information landscape within BIM.

2.10.1.3 Addressing Interdisciplinary Connections in BIM through Knowledge Graph Structures

Knowledge Graphs are instrumental in representing interdisciplinary connections within the BIM domain, enabling the synthesis of information across varied construction practices, regulatory requirements, and material specifications. This capability is particularly important as BIM increasingly intersects with fields such as environmental science, structural engineering, and sustainability studies. Through its graph-based structure, a KG can capture these interdisciplinary relationships, offering a scaffold for integrating emerging knowledge areas such as sustainability regulations and green building technologies within the existing BIM framework (Buehler, 2024). By revealing critical connections that influence design and construction processes, KGs help BIM practitioners make informed decisions based on a holistic view of relevant data sources, thereby fostering a more integrated approach to building management and project delivery.

2.10.2 Using TM in AEC industry

The application of TM in the AEC industry is gaining momentum due to its unique ability to extract valuable knowledge Weiss and uncover insights hidden within unstructured textual data (Aggarwal and Zhai, 2012; Weiss et al., 2010). As the AEC sector generates vast amounts of unstructured data, TM provides an effective approach to analyse this information, enabling project teams to make informed decisions and streamline project delivery processes. By utilising TM, AEC professionals can enhance project efficiency, optimise workflows, and better manage resources, ultimately contributing to improved project performance and timely delivery (Nayak et al., 2016).

Within the AEC industry, the adoption of data mining methods has not been widespread, but it has captured the attention of researchers to certain logistical considerations, such as the study of accident reports and the retrieval of reference events, provided that 80% of business information is in plain text, including the AEC industry (Ur-Rahman and Harding, 2012). Table 2.6 summarizes and compares recent literature that applied TM strategies to the AEC domains of corpus, methods, and results.

Table 2.6: Recent publications on TM applications in AEC

TM approach	**Corpus**	**Aim**
Rule-based topic model (Kim and Chi, 2019)	Accident reports	Quick retrieval of similar cases
Ensemble model to classify the causes of the accidents. Unsupervised chunking approach used to extract common objects (Zhang et al., 2019)		Technology validation identifying risky objects
Text classification (Goh and Ubeynarayana, 2017)		Technology validation of various algorithms
Count frequencies of 19 critical terms (Marzouk and Enaba, 2019)	Project contracts and correspondences	Data visualization (i.e., the correlation between different terms)
Count key terms; association rules of key terms (Hosseini et al., 2018)	BIM managers' job advertisements	Determine the role of BIM managers
Event detection (e.g., time, location, trigger and season) (Kim and Kim, 2018)	Fire-accident reports	Generate failure patterns
Clustering (i.e., tag allocation), and word cloud (Moon et al., 2018)	News reports of international construction projects	Easy retrieval
Count word pairs (Nedeljkovié and Kovaöevié, 2017)	Construction documents (e.g., change orders and weekly reports)	Better visualization of a key phrase network
Clustering (i.e., case-based reasoning) (Shen et al., 2017)	Buildings cases	Retrieve similar buildings cases for references in design
Information extraction (Yarmohammadi et al., 2017)	Software log texts	Transfer semi-structured data to structured, processed data
Text classification with ontology (Zhou and El-Gohary, 2017)	Regulatory documents	Classification of regulatory documents
Dictionary-based information extraction (e.g., injury precursors and outcome categories) (Tixier et al., 2016)	Injury reports	Transfer semi-structured data to structured, processed data
Entity recognition of BIM and GIS elements (Cheng et al., 2015)	IFC and City GML	Match terms with similarity measures
Count cost-related words (Williams and Gong, 2014)"	Short project summaries	Predict the level of cost overrun

The corpus varies from deals, accounts of incidents, logbooks to news stories. The techniques to TM are primarily used in the acquisition of information; identification of entities, core terminologies; grouping of documents; analysis of feeling; and subject models. It is possible to classify those features of the target items, for example, by using cost-related terms to estimate the degree or magnitude of how the incurred cost exceeded the project’s budget and relating to the incidence entities referred to in accident reports to explain the contributors of safety concerns.

Other estimation method types that can be used for classifying data include Bayes, decision tree, and ANN models. In general, the goals include converting unorganized data to organized or structured data, achieving better visualization of data, enhancing corpus classifications, and summarizing the pattern of information of certain variables such as emotions, perception and satisfaction. A BIM-cantered analysis of the insights extracted from the literature sources, as outlined in Table 2.4, includes a study conducted by Hosseini et al. (2018). In this study, a total of 199 BIM-related job advertisements were gathered from dedicated employment websites. The aim was to determine the frequency of keywords, thereby establishing a clear outline of the skills essential for a BIM manager role. The collection of studies detailed in Table 2.4 underscores the prospective utility of employing TM techniques to offer valuable insights into Architecture, Engineering, and Construction (AEC)-related concerns. These studies have thereby established a robust foundation for further research in this domain.

According to Ahmed et al. (2018), in the AEC domains, TM technologies are still in their infancy. As such, the topics are confined to some traditional concerns such as the review of accident reports, the retrieval of design reference cases, and the support for data interoperability through various software. Nonetheless, not much focus has been attributed to current issues encountered in BIM pervasive major project delivery. In addition, the corpus types depicted in Table 2.4, particularly incident reports, software log-book, and contracts, are semi-structured, and are generated by individual agencies. In this regard, it is easy to apply TM approaches to such data in comparison to unstructured data that construction players may come across in the field (Aggarwal and Zhai, 2012).

There are far more irregularities in public-generated unstructured data, such as user feedback, social media data, and user reviews; such data is not yet well-leveraged (Liu and Hu, 2019). Current research studies denote integrated domain expertise into TM procedures. A case in point is the work by Marzouk and Enaba (2019), their study involved compiling a set of 19 contract-related terms to facilitate the extraction of project contract data. Yet, the study highlighted that the incorporation of industry experience into this process has not been strongly emphasized.

Many studies do use ready-made methods, such as unsupervised topic models provided by Tencent, Amazon, and Google; and sentiment analysis engines (e.g., Text Blobs). Liu and Hu (2019) argued that even though the use of ready-made tools has surged, exploration of the viability and applicability these ready-made models is still lacking in terms of rigorous empirical works.

2.10.3 Assessing TM Adoption in Construction for Major Projects

Text/data mining in building projects is usually not common (Alsubaey et al., 2015). TM is aimed at extracting from unstructured text data or data that has not been subjected to analysis before. According to Williams and Gong (2014), the intensity of TM depends on the discovery of information concealed inside the text data. Miner (2012) identified five major forms of TM: retrieval of information, clustering of documents, classification of documents, extraction of information, and NLP.

Implementing TM techniques helps to eliminate cost overruns or rather surges in the cost incurred in delivering the project. To be precise, the stacking ensemble model can be used to predict the degree of cost volatility. Williams and Gong (2014) applied a textual data mining algorithm. The approach begins with the compilation of textual data that are the phrases that display the overview of the project and the words that define the project’s key mission. Subsequently, numerical data was obtained from text. Some of the obtained data pertained to dollar value sentences. As a final point, the two researchers concluded that it helps to avoid cost overruns by integrating numerical and text data. In order to extract details and coordinate documents, automated classification of construction paperwork has proved to be an integral TM technology (Caldas and Soibelman, 2003).

Ur-Rahman and Harding (2012) suggested textual mining techniques in other text categorization studies to promote market meaning in order to obtain vital information. First of all, the needless terms (we, is, are, of, a, an, etc.) are omitted to enforce the model. In addition, stemming is performed. Stemming can be contextually defined as a model through which words converge to their original source. Secondly, text data is interpreted in the form of a matrix, where the job is contained in the row vector, and the document ID is in the column vector. Thirdly, numerous algorithms have been used for paper grubbing. Finally, to differentiate post-project documentation into a positive or a negative document, several Main Term Phrasal Information approaches are adopted.

A predictive early warning technique for forecasting construction delays on other failures was developed by Alsubaey et al. (2015). Minutes of Meetings (MoMs) (unstructured data) were the principal records that were studied in the model. They began by selecting keywords from the available body of literature, such as: lack of labour force, lack of material on site, lack of overall safety, and many more. Subsequently, the researchers contrasted these keywords against MoM documents. The model leads to problems being found that may cause project failure. The presented model was focused on the recognition of keywords to acknowledge MoM sentiment.

Moreover, NLP has been implemented to evaluate and handle textual information pertaining to retrieving CAD sketches, undertaking automatic accident report analysis and finally automated similarity grouping of construction paperwork (Al Qady and Kandil, 2014; Yu and Hsu, 2013). In order to calculate the efficiency of the modellers, considering the time needed to execute similar operations, Yarmohammadi et al. (2017) applied TM to Autodesk Revit, using auto-generated log files with the aim of demystifying the modellers’ performance in reference to the time taken.

2.10.4 Summary

Studies carried out in the field of TM aim to achieve an optimal decision or to achieve a good project, and retrieve comparable historical cases. Applications in the fields of risk control, conflict resolution and green or environmentally sustainable building achievement have clustered distinct project paperwork or electronic documents into valuable classes, finally applying TM to calculate the productivity of employees. With regard to mining and the use of BIM in major construction projects, an exploration gap in the building sector is found to exist in this region. Few studies have uncovered the use of BIM in infrastructure, according to Shou et al. (2015). There is also a need for research to provide generalized knowledge.

2.11 Critical Appraisal of the Literature

In the systematic literature review conducted in this chapter, several methodological and knowledge gaps in the existing research on the use of TM in conjunction with BIM for performance enhancement in major construction project delivery were identified. These gaps highlight the need for further investigation and the potential contributions of the proposed research.

One of the methodological gaps identified is the limited application of advanced TM techniques. Many of the reviewed studies focused on basic methodologies such as keyword extraction and sentiment analysis, while advanced techniques like topic modelling, clustering analysis, finding associations, and entity recognition have received less attention. This indicates a need for research that explores the application of these advanced techniques in the context of BIM-enabled major construction project delivery. By incorporating advanced TM techniques, the proposed research aims to fill this methodological gap and provide a deeper understanding of how these techniques can enhance performance in BIM pervasive major project delivery.

Another methodological gap is the lack of standardized approaches for integrating TM with BIM. The reviewed literature lacked consensus on the steps and guidelines for implementing TM techniques in BIM-based project environments. To address this gap, the current research developed the BIMTAPE Framework, which provides a standardized approach for integrating TM in BIM pervasive major project delivery. This framework offers clear steps and guidelines for practitioners to follow, ensuring a systematic and structured implementation of TM techniques.

In terms of knowledge gaps, the existing research tends to overlook the specific integration of TM at different project phases, such as pre-construction, construction, and post-construction stages. Further investigation is needed to understand the unique contributions and benefits of TM in each project phase. The proposed research aims to address this knowledge gap by examining the specific application of TM techniques at different project phases and exploring how they can enhance performance and decision-making in each phase.

Additionally, while some studies acknowledge the importance of stakeholder analysis, there is a lack of research that explores the use of TM to extract insights from stakeholder feedback, social media data, or other textual sources related to project stakeholders. Understanding stakeholder perspectives through TM could provide valuable insights for decision-making processes and contribute to improved project performance. The current study research aims to fill this knowledge gap by examining the use of TM to analyse and interpret stakeholder-related texts and extract meaningful insights that can inform project management and enhance stakeholder engagement.

Overall, the identified methodological and knowledge gaps in the literature demonstrate the need for further research in the integration of TM with BIM for performance enhancement in major construction project delivery. This thesis addresses these gaps by developing the BIMTAPE Framework, exploring the specific application of TM techniques at different project phases, and investigating the use of TM for stakeholder analysis. By doing so, this research contributes to filling the identified gaps and advancing the understanding of how TM can improve performance in BIM pervasive major project delivery.

2.12 Summary

This chapter presents the systematic literature review using the NSP approach, and highlights key themes and areas related to TM for BIM pervasive major project delivery. The literature suggests that the use of BIM in major construction project delivery can have a positive impact and that TM can help to improve performance in BIM-focused major construction project delivery. TM can provide a new method to enhance the effectiveness of major construction project delivery. The BIMTAPE Framework is an example of how TM can support the implementation of BIM in major construction project delivery.

The insights gleaned from this chapter provide a solid foundation for the subsequent chapter, which delves into the research methodology (Chapter 3). The outlined themes and areas in the systematic literature review guide the approach taken in exploring the implementation and efficacy of TM for BIM in major project delivery. The research methodology section builds upon these insights to lay the groundwork for the investigative processes, ensuring a comprehensive understanding of the subject matter.

Chapter 3
Research Methodology

3.1 Introduction

This chapter outlines the comprehensive research methodology adopted to address the aim and objectives of this thesis, which seeks to enhance performance in major construction projects by integrating TM techniques within BIM. The research methodology employs a mixed-methods approach, using both qualitative and quantitative data derived from secondary sources. The methodology encompasses a systematic literature review, ontological modelling, and TM as key methods employed to extract valuable insights from unstructured textual data. Additionally, the validation of the research findings was conducted using Social Network Analysis (SNA), ensuring the reliability and credibility of the obtained results. In line with the research framework known as the "research onion" by Saunders et al. (2016), the chapter systematically presents the methodological choices, starting from philosophical assumptions to specific techniques for data collection and analysis. By employing these rigorous research methods, this study aims to contribute to the advancement of BIM practices and pave the way for improved project delivery outcomes.

The research employs a case study-based approach, grounded in Robert Yin's formal research methodology (Yin, 2014), which integrates both qualitative and quantitative paradigms. In accordance with Yin's framework, the study uses multiple sources of evidence and applies triangulation to ensure the validity and reliability of the findings (Yin, 2014). For the qualitative component, the research incorporates various methods such as case studies, literature reviews, and experimental approaches to gather in-depth data. These methods were chosen to provide a comprehensive understanding of the research problem, particularly within the context of BIM pervasive major project delivery. Yin’s methodology emphasizes the importance of case studies in exploring complex, real-world situations, such as the integration of BIM technologies with project management challenges (Yin, 2018).

The research integrates multiple methods within the qualitative and quantitative paradigms. While the qualitative aspect focuses on techniques like case studies, literature reviews, and experimental methods, the quantitative approach involves statistical analysis to complement and strengthen the qualitative findings. This combined approach allows for a thorough investigation into performance challenges related to budget, quality, and schedule in large-scale construction projects.

In line with Yin’s principles, the quantitative element uses statistical analysis to enhance the rigor of the research. By integrating both qualitative and quantitative methods, the study applies triangulation, which enables a more comprehensive examination of the research questions and performance challenges (Yin, 2006). This methodological integration, as guided by Yin's framework, ensures a robust and reliable analysis of the research problem (Yin, 1991).

Triangulation is a critical strategy in this research, reinforcing the credibility of the findings by employing multiple methods to investigate the same phenomena. This approach mitigates the limitations of relying on any single method, offering a richer, more understanding of the complexities involved in BIM implementation and its impact on construction project management (Yin, 2014). By leveraging multiple perspectives (Yin, 2018), the study gains deeper insights into performance improvements achievable through the integration of BIM and project management practices.

This chapter also elaborates on the limitations the author experienced during the data collection process. Study limitations and constraints are important aspects of any research that must be addressed and reported transparently. By including this information in the methodology chapter, the author allows the reader to understand the context in which the research was conducted and the limitations that may have affected the results. This helps the reader assess the validity and reliability of the findings and contributes to the transparency of the research process.

For data analysis, the research employed both qualitative and quantitative techniques, including non-statistical methods such as text analysis, and statistical methods. This mixed-method approach enabled the researcher to derive valuable insights from unstructured textual data, providing a detailed and comprehensive analysis of the issues at hand. The findings were further validated through SNA, which was utilised to enhance the credibility and reliability of the results, ensuring a refined and well-supported interpretation of the data.

Given the interdisciplinary nature of the research, which spans both construction and information technology, the pragmatic philosophical paradigm was selected as the most suitable. Pragmatism supports the use of multiple methods and data types, enabling the researcher to effectively address the research questions. Accordingly, a concurrent mixed-methods approach was adopted, allowing for the integration of qualitative and quantitative insights into the research findings.

Finally, this chapter discusses the limitations and constraints faced during the data collection process, offering transparency regarding the potential impact on the study's results. By acknowledging these limitations, the study provides a clear context for evaluating the validity and reliability of its findings. The chapter concludes by summarizing the adopted research methodology, which integrates multiple and mixed methods for advancing BIM practices and improving project delivery outcomes.

3.2 Methodology Based on Research Onion Model

Choosing a framework-based methodology is one of the challenging tasks in the thesis process. This is because every research has unique qualities and specifications. Several research methodology frameworks and models have been developed. One such strategy is the Research Onion Framework, proposed by Saunders et al. (2019). The research onion elaborates on various steps of writing a dissertation. The rationale behind using this framework is that it consists of a robust multidisciplinary approach. This framework has been used in previous studies related to the IT sector (see Karaseva, 2024; Meller, 2013). Several scholars from the IT sector have perceived the model feasible and suitable for defining the research methodologies. Saunders et al., (2019), the founders of this model, designed it in the following way (see Figure 3.1 Research Onion). The research onion model is supposed to be discussed from its outermost layer to the innermost layer.

Research paradigms shape how the world and its phenomena are perceived, understood, and interpreted. They represent the theoretical framework that guides the formulation and execution of research (Mackenzie and Knipe, 2006). Saunders et al.

(2019) define a paradigm as a method for gathering and interpreting knowledge about a given phenomenon. In the social sciences, which include construction management, paradigms are also known as theoretical perspectives (Crotty, 1998), research methodologies (Neuman, 2009), worldviews (Creswell, 2014), and systems of shared beliefs and practices (Morgan, 2007). Similar to the foundational elements of a building, Fellow and Liu (2008) argued that paradigms underpin the integrity of any research activity. Hall (2012) proposed that a worldview encompasses philosophical positions, including ontology, epistemology, axiology, and methodology. Therefore, paradigms can be distinguished through several interconnected aspects that define how knowledge is understood and approached in research (Haq, 2014; Williams, 2018):

- **Philosophy**: This concerns the origins and development of knowledge, guiding the overarching framework of inquiry.
- **Ontology**: Focused on beliefs about the nature of reality and what can be known, ontology shapes the assumptions researchers make about their subject matter.
- **Epistemology**: This explores the relationship between the researcher and the knowledge or reality being studied, influencing how knowledge is acquired and validated.
- **Axiology**: Concerned with the role of values in research, axiology addresses how a researcher's beliefs and biases may impact their work.
- **Methodologies**: These are the strategies and approaches used to access and generate knowledge, tailored to align with the philosophical and epistemological foundations of the paradigm.

These aspects together provide a comprehensive lens through which research can be structured and interpreted. In order to ensure that this study is framed with the appropriate perspectives, the remainder of this section will explore various elements of the research paradigm, including research philosophy, ontology, epistemology, axiology, methodology, and methods (Scotland, 2012).

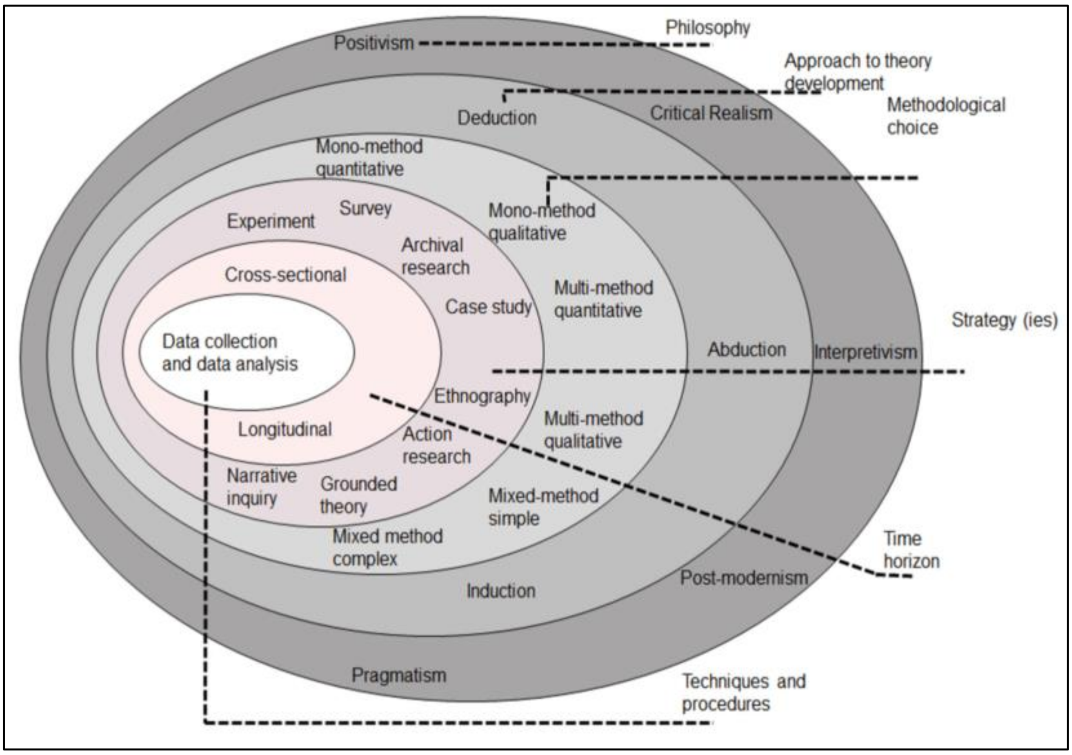


Figure 3.1: Research Onion Framework (Saunders et al., 2019, p.108)

3.3 Research Philosophy / Philosophical Paradigms

Prior to discussing the research philosophy, it is important to consider the different research modes elaborated by Gibbons et al. (1994) and Nowotny et al. (2003). These modes, known as Mode 1 and Mode 2, reflect distinct approaches to knowledge production. Mode 1 represents single-discipline research with clearly defined boundaries, while Mode 2, as defined by Nowotny et al. (2003), is "socially distributed, application-oriented, transdisciplinary, and subject to multiple accountabilities" (p. 179). The present research aligns with Mode 2, as shown in the Table 3.1 below:

Table 3.1: Different research modes

Mode 2 Characteristics	Relevance to the Research
Socially Distributed	While this research did not derive knowledge from a vast population, rich data was gathered from global scholars through an exhaustive literature review.
Trans-disciplinary	The research spans three primary disciplines: the construction sector, the IT sector, and project management.
Application-Oriented	The study's development of a robust framework and its application in data analysis demonstrates its practical, application-oriented approach.
Subject to Multiple Accountabilities	This research involves complex relationships with multiple stakeholders, including project managers, contractors, and IT experts, making it subject to various accountabilities.

Sources: Gibbons et al., 1994; Nowotny et al., 2003

According to Saunders et al. (2019), there are four key research philosophies: [1] positivism, [2] realism, [3] interpretivism, and [4] pragmatism. The present research adopts pragmatism, one of the central philosophies in the "research onion" framework. Pragmatism is particularly suited to this study, which employs a mixed-methods approach, integrating qualitative text analysis and quantitative methods such as SNA. To better understand the rationale for using pragmatism, it is essential to first review all the research philosophies briefly, considering their strengths and limitations, before justifying the selection of pragmatism as the most appropriate philosophy for this study.

3.3.1 Research Philosophy in Context

Research philosophy addresses the source, nature, and development of knowledge (Bajpai, 2011; Saunders et al., 2016). The philosophical assumptions made by the researcher guide the research process, influencing the choice of methods used to investigate the research problem. In social research, these assumptions are shaped by factors such as the existing body of knowledge and the nature of the problem being studied (Yin, 2003). The philosophical stance taken determines how knowledge is acquired and validated within a particular field. Resolving the question of research philosophy early on establishes a clear relationship between the researcher and participants, ensuring the adoption of appropriate methods for data sampling, collection, and analysis.

There are various models in the literature for understanding and classifying research philosophy in social and natural sciences (Guba and Lincoln, 2005; Tashakkori and Teddlie, 2003; Hall, 2013). However, according to the research onion model proposed by Saunders et al. (2016), the four main research philosophies - positivism, realism, interpretivism, and pragmatism - provide a foundation for the research. Each philosophy has its strengths and weaknesses, and the next step involves reviewing them to determine which philosophy is most appropriate for the study at hand.

3.3.1.1 Positivism

The positivist philosophy is based on the principle that reality is stable. It supports the scientific way of observing reality. Positivists believe that reality is objective and independent. While explaining various attributes of positivist philosophy, Phoenix et al. (2013) stated that positivism is the phenomenon of the Enlightenment Age that underpins the principles of natural sciences. It is a non-contextual, formal, and standardised research. Rationalism and empirical evidence are the bases of the positivist paradigm. Positivism is rooted in the principle called “a system of logic”, as Mill (1843) stated in his classical text. The present research cannot be called purely positivist research because it does not meet all of the following criteria of positivist philosophy (Park et al., 2023).

Dependent Variables: This research does not have dependent and independent variables for measuring as they exist in pure positivist philosophy.

Dualism: In the positivist paradigm, the researcher’s stand is objective. Separating from the participants is essential for minimising biases. However, this research does not allow the researcher to take only an objective stance, especially while synthesising the literature review.

Functional Relationship: It is an association between dependent and independent variables, which are measurable or quantifiable. Since the research does not have dependent and independent variables, there is no question of expressing the relationship between them using quantitative measures.

Hypothesis: It is a statement or idea that the theory is tested through quantitative or statistical tools. Technically, there should be two variables in the present research: BIM and TM and project management. The hypotheses should have been:

- ***H_a**: BIM and TM increase the construction project performance*
- ***H_0**: BIM and TM do not increase the construction project performance. (Null Hypothesis*

This research does not form such hypotheses and does not use statistical tests (e.g., the T-test) to test them. Therefore, though the research intends to observe the role of BIM and TM in construction project performance, the two variables are analysed through experiment methods to determine how they increase performance instead of whether they increase performance.

However, aligned with positivist philosophy, the research uses validity and reliability and SNA analysis, which is part of quantitative analysis. Thus, the present research follows a positivist approach, but along with positivism, it follows partial interpretivism.

3.3.1.2 Realism

Realism is the second philosophy suggested in the research onion. This philosophical school is based on the principle of mind-independent objectivity. It assumes that the external world is independent and objective, and it can be studied through observation and experience (Erikawati, 2023). Though, like realism, the present research follows rationalism and logical and critical thinking, it does not stop understanding the object, material or phenomenon (E.g., BIM, TM, and project performance). It does not observe but takes effort beyond just understanding the object with its reality. Therefore, realism is a one-dimensional approach to apply in the context of this research. Considering the attributes of realism, it is best suitable for descriptive research. However, the present research explores how to apply TM in BIM for project performance. Therefore, realism is not the appropriate philosophical stance.

The research onion mentioned interpretivism as one of the research philosophies. It is based on the subjective approach that social events or phenomena can be interpreted subjectively, and multiple interpretations can be possible. Interpretivism is a qualitative method in which data is collected from interviews, focus group discussions and observations by immersing oneself in the phenomenon. The present study has not used primary data, but the systematic review method is employed. It is a model-based study; therefore, the findings from the literature are used to design and apply the systematic model. TM and BIMTAPE Framework followed by experimental case studies. The findings are discussed descriptively, especially when the text analysis is performed qualitatively. Therefore, the research follows some inexplicit attributes of interpretivism.

3.3.1.3 Pragmatism

Pragmatism is the last philosophical stand suggested in the research onion. According to Pragmatism, some realities are not static but changing, and various factors influence reality. Pragmatist philosophers state that actions cannot be separated from situations (Morgan, 2014). Every individual's perspective and approach are different, and they try to find the reality by applying an independent approach. It also assumes that individual experiences are different. As a research paradigm, pragmatism believes that single or multiple realities are always open to several empirical inquiries (Creswell and Clark, 2011). Pragmatism is a very flexible approach, and it complements both quantitative and qualitative inquiries (Bryman, 2006; Teddlie and Tashakkori, 2003). The pragmatism approach also focuses on application value (BIM and TM application in improving project performance) rather than whether the outcomes are true (Does BIM and TM-based framework help to improve project performance). In this research, the focus is on the earlier outcome, i.e., focusing on application value.

This research indirectly follows the pragmatism approach. After critically reviewing all philosophical approaches from the research Onion, it was observed that pragmatism is close to experimental philosophy, an interdisciplinary approach in which two elements are primarily considered; first is the theoretical framework or model and the experiment methods (Knob and Nichols, 2017). Thus, the research is based on experimental pragmatism, in which the theoretical framework called TM and BIMTAPE is employed. Experimental pragmatism is used to test this model and its functional consequences. Empirical pragmatism is based on the tenet that scientific methods are essential to study philosophical ideas.

3.3.2 Comparison of Philosophies and Justification of Pragmatism

An overarching philosophical stance can provide a foundational belief system to a project aiming to conduct an original research inquiry. One assumption made from a philosophical perspective is that philosophical theory aimed at ontological and epistemological beliefs, as well as the methodology itself, must be conducted congruently (Goldkuhl, 2012). Positivism is the belief that statements reflecting observable facts align with the reality of the facts, and that the determinants of what the facts are, are independent of the values or beliefs of the researcher. The positive dimension of this philosophy is its emphasis on observable, physical reality and experiences. It also values a systematic approach to establishing the facts through explicit formal and stated methodologies that can be independently verified or tested.

Positivism maintains that whatever may be the object in point of subject, the scientific method only allows for quantifiable data. Research that follows this method collects a good deal of numerical or categorical statistical data, which is analyzed from an objective standpoint. Interpretivism, on the other hand, is unique in preferring a 'deep' understanding of the subjective nature of human behavior (Ryan, 2018). It is reluctant to lean on statistical data; it prefers the amorphous interpretative approach to understanding through the generation of propositions and hypotheses, consequently leading to building mid-range theories (Alharahsheh and Pius, 2020). Realism is a middle way between positivism and interpretivism, and some might say that, like pragmatism, it appears to be a philosophical compromise. It encourages us to ponder the polarities of HRM, for instance, practice and principle, employee perspective and employer viewpoint, reality and view, activity and perception, assumption and observer, scientific and religious, absolute and relative (Goldkuhl, 2012). One can draw significant polarities such as objectivity or subjectivity, interior and exterior viewpoints, idealism and realism.

Interpretivism has been labelled subjectivist or anti-positivist. Philosophers regard the anti-positivist label as inappropriate. Interpretivists seek an understanding of the phenomena in its richness of meaning from the subjective standpoint of human experience (Ryan, 2018). Realism has been used to refer to the idea of moving beyond mere reason or belief. For the purposes of this philosophical orientation, the focus is on social reality. The distinctions between positivist and interpretivist epistemology are characterized by binary opposites (Ryan, 2018). The philosophy of pragmatism is a fusion of positivism, anti-positivism, and realism that has gained favor in project management research, where it represents a 'broad church' within which several methodological perspectives are acknowledged. Pragmatism is an appropriate approach to project management research. Pragmatism is based not on theory but on the way in which it deals with its main problem, which is the problem in the conduct of human affairs. It is an attempt to effect a working compromise between radically different points of view about the ultimate nature of a philosophical theory (Goldkuhl, 2012).

After critically analyzing the philosophical positions proposed by Saunders et al. (2016), this study adopts pragmatism as its philosophical stance for several reasons. The study aims to explore how BIM and TM contribute to improving project performance and identify factors affecting their effective integration.

As suggested by Saunders et al. (2016) and Yin (2003), the choice of research philosophy should align with the nature of the research problem, research questions, existing knowledge, and available time for the study. The research problem involves the integration of BIM and TM, which are complex and context-specific, into construction project management. Pragmatism offers a pluralistic approach to address these challenges.

Pragmatism focuses on the "how" and "what" of research problems, based on the intended outcomes. The research questions of this study require both "how" and "what" to be addressed. As stated in Chapter 1. Answering these questions requires a flexible worldview that embraces multiple research philosophies (Creswell, 2014). Pragmatism allows researchers the freedom to choose the most appropriate methods, techniques, and procedures for addressing the research problem. As pragmatism supports both positivism and interpretivism (Parvaiz et al., 2016), this pluralistic approach was adopted as it addresses the research questions in a way that ensures positive practical outcomes (Teddlie and Tashakkori, 2010). Furthermore, pragmatism is well-suited to problem-solving in human activities—here, how BIM and TM can facilitate better decision-making and performance in construction projects.

3.3.3 Ontological and Epistemological Assumptions of the Study

As specified earlier, this research is a mixed method. Therefore, ontology and epistemology are the two philosophical foundations that guide the researcher in applying mixed methods to the study. Therefore, it is essential to understand the terms ontology and epistemology and their significance in the present research. Ontology has been precisely defined in the previous studies. For example, Gray (2013) defines ontology as the viewpoints about the nature of entities and various assumptions about them (Bryman and Bell, 2011). There are two ontological perceptions. According to the first assumption, reality is one and the second perception is that multiple realities are there in the world. The first viewpoint or perception motivates the research to be objective, whereas the second viewpoint is inclined toward subjectivity. The two are polar opposite perceptions. The researcher's ontological viewpoints used for this research can be positioned between a single reality and multiple realities. It can also be called the intermediate position. The following Table 3.2 elaborates on how two perspectives (first and second viewpoints of reality) are used.

Table 3.2: The researcher's ontological viewpoints

Single Reality	**Multiple Realities**
Frequency Analysis	Word Cloud
Trend Analysis (Graph Results)	Co-occurrence Network Keywords
Publication Timetables	Network of Authors Contributing to BIM Research and Trends
Country-wise Publication Activities	Global Distribution of Published BIM Studies
Formula for Phrase Model	Codings Bibliometric Analysis

Like the ontological philosophy, the researcher's stand is objective while analysing measurable variables. Here, the main focus is on the evidence. While holding the epistemological approach, the data is understood and coded using a subjective approach.

The intermediate viewpoint on epistemology with positivist and interpretivist approaches helped the researcher study empirically how BIM and TM help to improve project performance and how effectively they can be integrated into the construction sector.

3.4 Research Approaches for This Study

Research Onion has three research approaches to theory development: Inductive, deductive, and abductive research approaches. The study adopts an abductive research approach. Like Saunders et al. (2019), many scholars have focused on inductive and deductive approaches. Trochim (2006) also discussed inductive and deductive approaches. The inductive approach goes from specific to general, and the deductive approach goes from general to specific. Cresswell and Plano Clark (2007) defined the deductive approach as “a top-down” approach, “from a theory to hypotheses to data to add to or contradict the theory” (p.23).

In contrast to this, the inductive approach takes a "bottom-up" perspective, relying on participants’ views to construct broader themes and develop a theory that interconnects these themes (Cresswell and Plalno Clark, 2007, p.23). The deductive research approach is used in quantitative analysis, whereas the inductive approach is used in qualitative analysis. Traditionally speaking, while applying the deductive approach, the researcher first develops hypotheses and then tests them to derive inferences. Statistical methods are used to test these hypotheses. Trochim (2006) further argues that qualitative research is inductive and exploratory, while quantitative research is deductive and confirmatory, in which relationships between different variables are tested.

Abduction is the third tradition of logical reasoning, alongside deductive and inductive logic, and a recent addition to academic theorizing. However, its development has strong links with some of the earliest traditions of philosophy, in particular the phenomenology of pre-Socratic philosophers. Like the traditional deductive method of moving from the general to the particular (and thus, in truth, theoretical), abduction is about the discovery of, and moving from, particular instances to something more general (and thus practical) (Okoli, 2023). Unlike these foundational thinkers, abduction has been theorized paradoxically both as a form of reasoning and as a basis for a new, pragmatic theory of change. This paradox is reflected in the methodological literature as the abductive approach is debated as both a means to explore unknown contexts and develop new theories and problem-solving pragmatically in a real-world context.

3.4.1 Deductive Research Approach

In essence, deductive reasoning is a process of testing theoretical propositions derived from existing theories or frameworks against empirical data to determine the relationship between two different variables within the theory for the purpose of either falsifying or revising the theory or generalizing the theory to a new population (Okoli, 2023). When employed, the deductive research approach is marked by a systematic, targeted, and design-driven methodology and involves a structured process. If the hypotheses, data, and interpretations are valid, the theory is justified. Knowledge generated by this method is often theoretical, premised on logic and abstraction. In practice, both qualitative and quantitative data can be analyzed using the deductive approach to address the stated research question. In deductive research, scholars generally look for generalizable results (Osman, et al., 2018).

Deductive research has some distinct strengths. By operationalizing theoretical knowledge and testing hypotheses based on the adequacy of available theories, researchers get a good sense of the direction to pursue the research. Further, it is relatively easy to falsify deductive research, and the results can be validated before employing broader research (Osman, et al., 2018). Yet such research is inherently weakened by the fact that researchers may not be able to discern hypotheses from adequate theories and reasonable assumptions, and the findings sometimes leave out unique project characteristics so as to make the results generalizable. Although the deductive research approach can help project management build a theoretical knowledge base, it can also help to identify the limitations of existing theories for future research. This identifies how current and past researchers have defined the terms and conditions or interpreted the relationships among the causal variables, thereby revealing the risk of theoretical and econometric artifacts (Osman, et al., 2018). The inconclusive studies could create clarity of reason to move into customized qualitative research.

3.4.2 Inductive Research Approach

The inductive research approach seeks to generalize specific observations and link them to general statements, which is the opposite of the deductive research approach. The deductive approach is first hypothesized and then verified using analysis (Okoli, 2023). Its main characteristic is the exploration or search for new theories, as its objective is to generate concepts and hypotheses for later testing by means of collecting new data that has not been seen before. This type of research has flexibility, openness, and responsiveness to the surrounding context, as well as the ability to understand the world without relying on certain assumptions.

Inductive reasoning is a logical approach that involves formulating general principles from specific instances (Osman, et al., 2018). It moves from a narrowly focused set of observations to broader generalizations. It is precise and direct, uncovering patterns and insights that other research methods cannot match. It can capture unique thoughts and ideas using various methods. While deductive and inductive reasoning have theoretical principles, an inductive approach takes a more open-ended research process. Additionally, an inductive approach can be useful in project management because the research field is enriched with a wide variety of research methods that deal with the research context and its dynamics. An inductive approach to research is utilized when data is collected and the research variables are not well structured. Furthermore, an inductive approach to research is best suited for research that is exploratory in nature (Okoli, 2023).

3.4.3 Abductive Research Approach

Abduction, also known as the retrodictive approach, is intended to bridge the gaps left by inductive and deductive approaches by adopting a pragmatic stance. As suggested by Saunders et al. (2012), the deductive approach can lack clarity in how to select a theory to test, while the inductive approach is criticized for its inability to build theories purely from empirical data. The abductive approach overcomes these limitations by focusing on "incomplete observations," "surprising facts," or "puzzles," which are addressed using a combination of qualitative and quantitative methods (Bryman and Bell, 2015).

Given the aim of this study, to develop a framework for using BIM and TM in project performance, the abductive approach is considered suitable. This approach allows for the iterative use of both deductive and inductive reasoning, providing a comprehensive way to address the research questions. The abductive approach enables the researcher to develop theories based on unexpected findings and test those theories using subsequent data collection. Table 3.3 presents a comparison of the three research approaches: Deductive, Inductive, and Abductive.

Table 3.3: Comparison of the Three Research Approaches

Aspect	**Deductive Approach**	**Inductive Approach**	**Abductive Approach**
Logic	In a deductive inference, when the premises are true, the conclusion must be true.	In an inductive inference, known premises are used to generate untested conclusions.	In an abductive inference, known premises are used to generate testable conclusions.
Generalisation	Generalizing from the general to the specific.	Generalizing from the specific to the general.	Generalizing from the interactions between the specific and the general.
Use of Data	Data collection is used to evaluate propositions or hypotheses related to an existing theory.	Data collection is used to explore a phenomenon, identify themes, and create a conceptual framework.	Data collection is used to explore a phenomenon, identify themes and patterns, locate these in a conceptual framework, and test this through subsequent data collection.
Theory	Theory falsification or verification.	Theory generation and building.	Theory generation or modification, incorporating existing theory where appropriate to build new or modify existing theory.

Source: Adapted from Saunders et al. (2016)

3.4.4 Justification for the Choice of Research Approach

In the project management field, researchers are continuously looking for ways to improve how the performance of projects can be enhanced, transformed, and conducted through the lens of different methodologies, methods, and techniques. The researcher argues that a research strategy based on the examination of ambiguous relationships utilizing an abductive approach can potentially lead to exciting and new research insights and outputs that are set to shape the future of modern project management research (Osman, et al., 2018). More importantly, this study believes that the justifiable need for additional research on the language and logic of these methods leads to a compelling case for why a pragmatic abductive reasoning approach should be utilised in project management research, employing an exploratory strategy that seeks to create research outcomes relevant for understanding the dynamics of a complex and changing world. The abductive approach is useful for generating new discussions and arguments for problem-solving in our real-world practice and improving the theoretical construction of project management. Given that every project can be affected by a host of factors including cultural differences, psychological, social, and economic industrial contexts, abductive reasoning and qualitative research methods are needed to uncover the complexities underpinning project actions and to make any practical or theoretical contributions to the quality of project outcomes (Osman, et al., 2018).

In addition, after critically analyzing the available philosophical positions proposed by Saunders et al., (2016), this research adopts pragmatism as its philosophical stance for several reasons. This study aims to explore the role of TM and BIM in improving project performance and to identify and measure the factors impacting their effective integration. According to Saunders et al. (2016) and Yin (2003), the choice of research philosophy should be guided by the nature of the research problem, the type of research questions, existing knowledge within the field, and the time available for the study. In this research, the problem involves the integration of TM and BIM into major construction project management, focusing on identifying factors influencing their effective use. Pragmatism provides the necessary pluralistic approach to address the complex nature of these research questions.

This research adopts the abductive research approach because it best addresses the research aim and objectives and aligns with the research paradigm and philosophical stance already adopted. The abductive approach allows for the iterative development of a framework that integrates TM and BIM, enabling the researcher to explore and explain research constructs and variables.

Accordingly, the study began with a systematic literature review (similar to archival research) to develop a preliminary conceptual framework (deductive approach). This framework was then refined through experiments and case studies. These experiments, treated as case studies, involved the analysis of academic articles, policy papers, and technical documents, allowing the researcher to gather qualitative and quantitative data to explore the role of TM and BIM in project performance enhancement.

3.5 Methodological Choice

The selection of research methodology forms the third layer of the 'research onion' model, emphasizing the plan of action or design that guides the choice of techniques applied in the research process (Crotty, 2003). This layer outlines the philosophical framework that supports the research methods, addressing the rationale for selecting either qualitative, quantitative, or a combination of both methods. Saunders et al., (2016) outline three primary categories of research methodologies: mono-methods (qualitative or quantitative), multi-methods (qualitative and quantitative), and mixed-methods (simple and complex). Mono-methods pertain to the exclusive use of either qualitative or quantitative methods, while multi-methods and mixed-methods involve the integration of both qualitative and quantitative approaches to explore a single research inquiry.

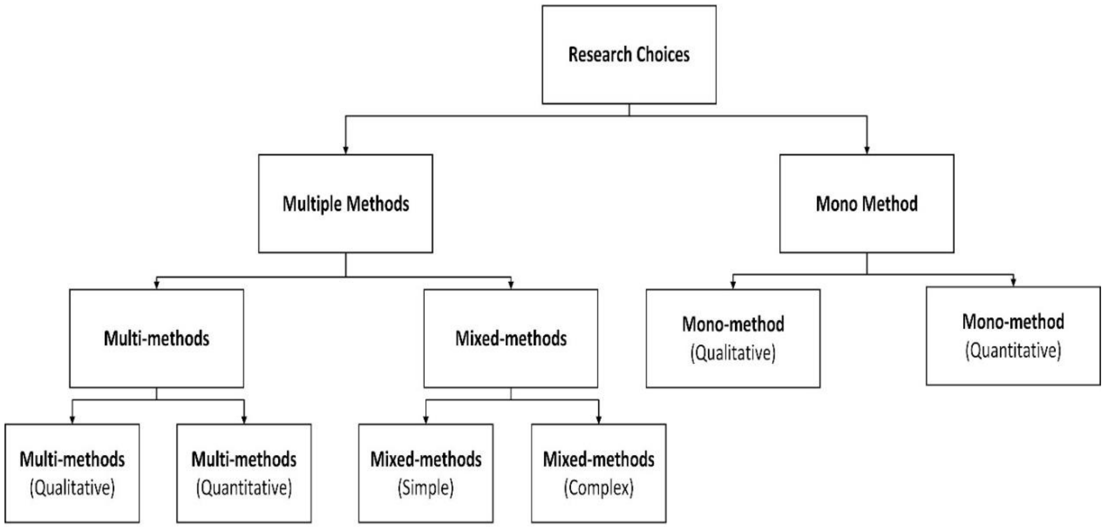


Figure 3.2: Research methodological choices.
Source: Saunders et al. (2009)

Saunders et al. (2019) focus on two primary methods: quantitative and qualitative. In their previous model, the authors suggested methodological choices are mono, mixed and multi-methods. But later in the revised model, they further segregated these three methods:

- Mono method qualitative
- Mono method quantitative
- Multi-method qualitative
- Multi-method quantitative
- Mixed method simple
- Mixed Method complex

Here, three core or primary methods are considered, Table 3.4.

Table 3.4: Three core or primary methods

Mono Method	**Multi-Method**	**Mixed-Method**
In the mono method, any one method (either qualitative or quantitative) is used. Saunders et al., (2019) segregate it into two parts: Mono method quantitative, and mono method qualitative	When the researcher uses a single method (e.g., qualitative or quantitative), multiple research tools are used within the same method. For example, if the researcher is conducting a qualitative study, he or she can use two or more tools, such as interviews, focus groups, observation and case studies.	Mixed-method is the combination of qualitative and quantitative methods. For example, the researcher uses a survey questionnaire followed by statistical analysis (as a quantitative method) and in-depth interviews and focus group discussions followed by thematic, content, narrative, discourse or other such analyses (quantitative method).

However, this research is a bit complex. It uses multiple and mixed methods. While applying the mixed method, the researcher has used quantitative and qualitative data from secondary sources. For analysing the data derived from both methods, the researcher uses qualitative and quantitative data analysis methods (non-statistical, text analysis, and statistical). Further in the qualitative method also, the researcher used multiple methods to collect the data, case studies, literature review, and experiments.

It is a multi-method because it applies statistical analysis experiments (Quantitative method) and systematic review and text analysis (qualitative method). Therefore, the methodological choice of this study is mixed-complex.

3.5.1 Mono-method (Qualitative and Quantitative) Research Methodology

The mono-method research methodological choice refers to a scenario where the researcher investigates a social or human phenomenon using a single method. This method could either be qualitative (e.g., unstructured interviews) with qualitative data analysis, or quantitative (e.g., questionnaires) with quantitative data analysis (Saunders et al., 2019). These scenarios are referred to as mono-method qualitative and mono-method quantitative, respectively. This approach was not adopted in the current study due to the need for integrating multiple tools and methods for a comprehensive investigation.

3.5.2 Multi-method (Qualitative and Quantitative) Research Methodology

Multi-method refers to a research design where more than one data collection technique is used within a single research method, either qualitative or quantitative (Tashakkori and Teddlie, 2003). There are two scenarios:

(1) Multi-method qualitative: Collecting textual data (e.g., through semi-structured interviews and diary accounts) and analyzing it using qualitative procedures (e.g., content analysis).

(2) Multi-method quantitative: Collecting numerical data using multiple techniques (e.g., questionnaires and experiments) and analyzing it using statistical methods (Saunders et al., 2012).

In multi-method approaches, qualitative and quantitative techniques are not mixed but used within the same overall methodology. In this research, multiple qualitative and quantitative tools were employed to collect and analyze data.

3.5.3 Mixed Methods (Simple and Complex) Research Methodology

Mixed methods research involves the combination of qualitative and quantitative data collection techniques and analysis procedures within a single study. There are two forms:

- Mixed-methods simple: This involves using qualitative and quantitative techniques concurrently (in parallel) or sequentially, without combining the results. Qualitative data are analyzed qualitatively, and quantitative data are analyzed quantitatively.
- Mixed-methods complex: In this approach, qualitative and quantitative data collection techniques and analysis procedures are combined at different phases of the research. For example, qualitative data might be quantitised, and quantitative data might be qualitised (Saunders et al., 2009).

According to Creswell (2014), mixed methods can be further classified into:

(1) Convergent Parallel Mixed Methods: Quantitative and qualitative data are collected and analyzed concurrently.
(2) Explanatory Sequential Mixed Methods: Quantitative data is collected and analyzed first, followed by qualitative data.
(3) Exploratory Sequential Mixed Methods: Qualitative data is collected and analyzed first, followed by quantitative data.
(4) Transformative Designs: These designs use either qualitative or quantitative methods, but within a transformative framework or lens.

Given the complexity of this study, mixed-complex methods were employed. Data was collected and analyzed both concurrently and sequentially to ensure a comprehensive understanding of the research questions.

3.5.3.1 Justification for Mixed-Complex Methods Choice

The choice of mixed-complex methods was deemed essential due to the multi-faceted nature of the research questions and objectives. This approach provided several benefits, including:

- Triangulation: Using multiple independent data sources or collection methods to corroborate findings.
- Facilitation: Using one method or strategy to assist another within the study.
- Complementarity: Employing multiple strategies to investigate different aspects of the research.

- Aid in interpretation: Qualitative data helps to explain relationships between quantitative variables.
- Generalization: This approach allowed the findings to be contextualized and generalized to a broader scope.

By utilizing multiple methods, this study aimed to explore both qualitative and quantitative aspects of the research questions. The integration of tools such as systematic literature reviews, case studies, text analysis, and experiments ensured robust data collection and analysis, yielding more reliable results and conclusions.

Table 3.5: Reasons for Using Mixed Methods Designs in Research

Reason	**Explanation**
Triangulation	Using two or more independent data sources or data collection methods to corroborate findings.
Facilitation	Using one data collection method or research strategy to aid another within a single study.
Complementarity	Using two or more research strategies to dovetail different aspects of the research investigation.
Generality	Contextualizing the main study or using quantitative analysis to provide a sense of relative importance.
Aid Interpretation	Using qualitative data to explain relationships between quantitative variables.
Study different aspects	Using quantitative methods for macro-level study and qualitative methods for micro-level aspects.
Solving a Puzzle	Using an alternative method when initial methods are insufficient or ambiguous.

Source: Saunders et al., (2016)

3.6 Research Strategies

The Research Onion model includes seven fundamental research strategies. In line with Saunders et al., (2016), several strategies can be employed depending on the research objectives, questions, and data sources. These strategies include experiments, surveys, case studies, action research, grounded theory, ethnography, and archival research. Each strategy has unique characteristics suited to different types of studies.

The following Table 3.6 presents various research strategies and their characteristics as outlined by Saunders et al., (2016). The thumb rule is that surveys and experiments are used in a quantitative approach, whereas ethnography, grounded theory, action research, and archival research are qualitative studies. The case study method is used in both quantitative and qualitative research. However, it is not strictly defined, and the approaches and strategies can overlap according to the specific demands of the research. For example, surveys and experiments can be qualitative, whereas case studies can be both quantitative and qualitative.

Table 3.6: Research Strategies and their Characteristics

Research Strategy	**Characteristics**
Experiment	• Suitable for laboratory research rather than the field • Unlikely to be related to the real world of organization
Survey	• Used for exploratory and descriptive research • Most frequently used to answer what, who, where, how much, and how many questions • Easy to explain and to understand research strategy
Archival research	• Make use of administrative records and documents as the principal source of data • Allows research questions which focus upon the past and changes over time to be answered
Case study	• Suitable for research which wishes to gain rich understanding of the research context and processes • Able to generate answers to the research questions why, what, and how • Not suitable for collecting that data for generalisation
Action research	• Provides an in-depth understanding of specific phenomena • However, literature advises using it mainly in the education context
Ethnography	• Used to study groups of people • Requires a longer-term of fieldwork study
Grounded theory	• Data collection process might require several field visits • Criticised for its confusing process and time required for completion
Narrative inquiry	• Suitable for small, purposive samples • Intensive and time consuming
Phenomenology	• Suitable for investigating participants' worldview and experiences with respect to a phenomenon • Concerned with individual's perception of the phenomenon under investigation

Source: Saunders et al. (2016)

3.6.1 Research Strategy Adopted in This Study

In this research, a combination of experimental case studies and systematic literature review was adopted as the primary research strategies. The research conducted six experiments in combination with case studies, Table 3.7.

Table 3.7: Experiments in combination with case studies

Experiments	**Cases**
Experiment 1 – BIM-TS 1	Academic Articles
Experiment 2 – BIM-TS 2	Technical Files from Case Projects
Experiment 3 – BIM-TS 3	Approved Documents
Experiment 4 – BIM-TS 4	British Standard TM Process
Experiment 5 – BIM-TS 5	BIM Policies at Companies
Experiment 6 – BIM-TS 6	National Policy Papers

The TRIZ (Theory of Inventive Problem Solving) and Evidence-Based Learning (EBL) methods were integrated to establish preliminary findings. TRIZ was used as a systematic literature review tool, which justified the role of TM for BIM-integrated project delivery. Ontological modelling was then employed to establish the new framework, addressing the second research objective. Finally, TM experiments were conducted to verify the utility of the created framework, fulfilling the third research objective. These experiments reflected the current state of research and best practices in major construction project environments.

The combination of experiments and case studies was employed under a strategy termed experimental case studies, where the researcher reviewed policy papers, articles, and other technical documents, and conducted experiments, which were treated as case studies. The data obtained through academic articles, policy papers, and reports were analyzed using BIM, TM, and relevant models, forming a robust foundation for the research.

The adoption of TRIZ as a systematic literature review method proved to be valuable in identifying problems and generating creative solutions, which is critical for overcoming challenges in TM usage for BIM pervasive major construction projects (Gadd, 2011). Furthermore, the NSP approach was used to structure the systematic literature review, enabling the researcher to explore the history, present, and future of TM and BIM in major construction projects. The NSP facilitated a qualitative identification of themes, which informed the BIMTAPE Framework for project management, addressing the study's third research objective.

By combining systematic literature review, TM experiments, and case studies, the study adopted a mixed experimental case study strategy to ensure comprehensive data analysis. This approach proved effective in identifying research themes and gaps, guiding the research roadmap for the TM approach for BIM-oriented performance enhancement in major construction projects.

Thus, the abductive approach allowed for the iterative combination of both inductive and deductive reasoning, which aligns with the overall research aims. This approach facilitated a back-and-forth process between these methods, enabling the development and refinement of the BIMTAPE Framework. Furthermore, the abductive reasoning process aligns with the pragmatist stance already adopted in this study (Parvaiz et al., 2016; Morgan, 2007). This iterative process ensured that the research remained flexible, responding to emerging insights from both theory and practice.

This study's approach is highly suited to complex, non-traditional research frameworks, focusing on data from academic literature, policy documents, and technical files rather than primary sources like surveys and interviews. The complexity of the research and the integrated use of various frameworks and software tools justify the adoption of the mixed experimental case study strategy.

3.6.2 Justification for Research Strategy

The combination of experimental case studies and systematic literature review was selected to thoroughly investigate the integration of BIM and TM in major construction projects. This approach allowed for:

(1) Rich Contextual Understanding:

The case study approach, with a focus on real-world applications of BIM and TM, provided a deep understanding of how these methodologies influence project performance. Each case was an opportunity to explore how the theoretical concepts are applied and adapted in practice, offering insights into their functional consequences.

(2) **Empirical Testing through Experiments:**

The experimental nature of the case studies enabled the testing of the BIMTAPE framework under controlled yet realistic conditions, ensuring that the findings were not only theoretically sound but also applicable in practical construction project environments.

(3) **Thematic Exploration and Gap Identification:**

The systematic literature review allowed the researcher to critically assess existing literature on BIM and TM, informing the framework's development and ensuring that it addressed contemporary challenges and filled gaps in the literature.

3.6.3 Justification for Multiple and Mixed Methods

Given the complexity of this research, a multiple and mixed-method approach was essential for providing a comprehensive analysis of both qualitative and quantitative data. The following justifications explain why this approach was selected:

(1) **Triangulation:** The combination of multiple data sources (systematic literature review, case studies, and experiments) enhanced the validity of the research by corroborating findings from different methods.

(2) **Complementarity:** The use of both qualitative and quantitative techniques allowed for a more thorough exploration of the research questions. Qualitative insights from case studies were complemented by the quantitative analysis derived from experimental data, leading to a richer understanding of the integration of BIM and TM.

(3) **Facilitation:** The experiments conducted within the case studies were guided by insights gained from the systematic literature review, demonstrating how one research method supported and informed the other.

(4) **Generalisation and Interpretation:** While the case studies provided detailed, context-specific insights, the combination of experiments allowed for the generalisation of findings. This mixed approach also facilitated the

interpretation of complex relationships between BIM, TM, and project performance.

This comprehensive strategy enabled a robust and reliable exploration of the research problem, ensuring that the BIMTAPE framework was well-founded both theoretically and practically.

3.7 Time Horizon

Conducting research is a time-intensive process that depends on the duration spent on data collection and analysis. The fifth layer of the ‘research onion’ represents the time horizon, which refers to the period within which research is conducted to answer the research question (Saunders et al., 2012). Saunders et al., (2019) categorize research based on two types of time horizons: cross-sectional and longitudinal. Additionally, Kosow and Gaßner (2008) identified three fundamental time horizons:

Table 3.8: Time horizons

Short-Term	**Medium Term**	**Long-term**
Up to 10 years	Up to 25 years	More than 25 years.

Cross-sectional research, also referred to as "one-shot research," involves data collection at a single point in time. It is typically shorter in duration and less resource-intensive than longitudinal research, which collects data over an extended period. In longitudinal studies, data is collected at different points in time, often over years, allowing researchers to compare data sets and study changes over time. Cross-sectional studies are used when time or resources are constrained.

The present research adopts a cross-sectional time horizon, as data was collected from articles, reports, and policy papers at a single point in time. This one-time data collection is consistent with the limitations of cross-sectional studies, making the research relevant only for the current timeframe. As technology evolves, the findings and frameworks established in this research may become outdated, providing future researchers with opportunities to propose updated models that align with new technological advancements and circumstances.

3.8 Data Collection and Analysis

The collected data underwent qualitative TM techniques to extract meaningful information. This involved the use of NLP algorithms to analyse the text data. The qualitative TM process included steps such as data pre-processing, keyword extraction, entity recognition, and topic modelling. These techniques aimed to uncover patterns, themes, and relationships within the text data.

To analyse the results obtained from qualitative TM, the researcher employed quantitative analysis techniques, specifically SNA, a methodological approach used to analyse relationships and interactions among entities within a system. In this research, SNA was applied to the extracted data to examine the relationships and connections between different keywords, concepts, or entities identified through TM.

The application of SNA allowed for the visualisation and analysis of network structures and patterns present in the data. It provided insights into the centrality, connectivity, and influence of keywords or concepts within the BIM pervasive major project delivery domain. Through quantitative analysis using SNA, the researcher aimed to uncover hidden relationships, identify influential factors, and gain a deeper understanding of the interactions between different elements.

It is the innermost layer of the Research Onion model. The philosophical stance, research approach, and methodological choice determine data collection and analysis methods. This research does not contain the human participants. Furthermore, it is not a conventional experiment method in which a control group or experimental group. However, the experiment method is employed using secondary data such as documents, reports, policy papers and journal articles. The unique characteristic of this research is that a single article or document is used as a case study separately for every experiment. These documents were thoroughly studied to extract quantitative and qualitative data relevant to the research.

The basic statistical methods, frequency analysis and graphical presentation of the data, are applied while analysing quantitative data.

In qualitative analysis, the research employed the content analysis method. Following are some of the types of content analysis.

- Relational analysis
- Conceptual analysis
- Proximity analysis
- Cognitive mapping

Among these types, proximity analysis was perceived to be the most suitable method for the present study. Proximity analysis is the process of evaluating the existence of the explicit concepts present in the text. The text in the research is defined as a string of words called a window. Proximity analysis is the chosen method for a document-term matrix in which each row represents a document, and a column represents a term.

While collecting and analysing the data, the researcher strictly followed some protocols, as research is a publicly accountable process. TM and analysis consists of big data access, which is public or private data. The concept of privacy is very crucial in TM. While collecting and analysing data, the data privacy protocols were strictly followed. The researcher confirmed that the documents, reports, policy papers or articles were credible and reliable and not opinion-based random articles or content. The reliability and credibility of the data were double-checked.

As suggested by Wilkinson et al. (2016), the data analysis is performed using the FAIR principle, which stands for findable, accessible, interoperable, and reusable. The guiding principles are as follows, Table 3.9:

Table 3.9: FAIR principles of data analysis

Findable F1. (meta)data are assigned a globally unique and persistent identifier F2. data are described with rich metadata (defined by R1 below) F3. metadata clearly and explicitly includes the identifier of the data it describes F4. (meta)data are registered or indexed in a searchable resource	**Accessible** A1. (meta)data are retrievable by their identifier using a standardised communications protocol. A1.1 The protocol is open, free, and universally implementable A1.2 the protocol allows for an authentication and authorisation procedure, where necessary A2. Metadata is accessible, even when the data are no longer available
Interoperable I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation. I2. (meta)data use vocabularies that follow FAIR principles I3. (meta)data include qualified references to other (meta)data	**Reusable** R1. Meta (data) are richly described with a plurality of accurate and relevant attributes R1.1. (meta)data are released with a clear and accessible data usage license R1.2. (meta)data are associated with detailed provenance R1.3. (meta)data meet domain-relevant community standards

Source: Wilkinson et al. (2016)

3.8.1 Research Techniques and Procedures for the Study

This research was based on the three objectives described in Chapter 1. The intention was to ensure that the objectives were met by adopting the best method for the study that would guarantee positive outcomes. Accordingly, the focus was placed on leveraging the systematic literature review to inform the study on the use of TM for BIM-pervasive major construction projects and the need for the development of new technical solutions as an approach to boost performance-oriented practice in major construction projects. The research was conducted by considering the three objectives described in Chapter 1.

The research methodology adopted in this study follows the research onion model (Saunders et al., 2016), providing a structured framework that moves from philosophical considerations to data collection and analysis techniques. A convergent parallel mixed-methods approach was selected to address the research questions. This approach integrates qualitative TM, quantitative social network analysis (SNA), and a SLR enabled by the NSP. The SLR was employed to review and synthesize existing literature, while TM was used to process and extract patterns from unstructured textual data (Hotho, Nürnberger and Paaß, 2005), and SNA provided quantitative validation by mapping relationships between key terms and themes.

3.8.1.1 Systematic Literature Review with Nine-Square Process

The SLR employed in this research was based on the NSP, a structured framework for organizing and analyzing a wide body of knowledge. The NSP divides the literature into nine distinct categories or dimensions, providing a comprehensive and systematic approach to organizing and analyzing existing studies (Gadd, 2011). By applying the NSP, this study was able to:

- Ensure a holistic review of the literature related to BIM and TM.
- Minimize biases by categorizing sources into distinct themes, such as BIM performance, TM applications, and social network analysis.
- Identify research gaps and formulate a solid theoretical foundation for the empirical work.

This systematic approach ensured that the literature review was both comprehensive and structured, as each category of the NSP was carefully analyzed. The insights gained from this review informed the development of the research methodology, ensuring that the chosen methods were well-supported by the existing body of knowledge.

The use of the TRIZ as a systematic literature review method is a valuable approach in research. TRIZ is known for its ability to help identify problems and find creative solutions (Gadd, 2011), which can be useful in uncovering the challenges facing the use of TM in BIM pervasive major construction projects. The NSP is a structured approach that allows the author to examine the various challenges and potential benefits associated with TM in this context (Chen et al., 2021). In this study, the NSP is critical in uncovering the potential benefits of TM, such as minimizing time and material waste and delivering maximum value to the client. This information can then be used to inform the development of a new framework for TM in BIM pervasive major construction projects, as well as to validate the usefulness of the framework through TM experiments.

TRIZ is an abbreviation of the Russian term “teoriya resheniya izobretatelskikh zadatch”, translated into English as “theory of inventive problem-solving tool”. It was originally invented by Genrich Altshuller in the Soviet Union in 1946. Since then, it has gained wide adoption in industry including in the built environment sector. It has been found to be useful in establishing a comprehensive understanding thematic problems (Gadd, 2011), and it was accordingly adopted as a research method in the study to identify themes and inform the roadmap of this study.

Chen et al. (2007) adopted NSP and ANP in performing a multi-criteria assessment of façade systems while considering the useful life of the design. In the study, NSP facilitated a comprehensive systematic literature review that established a technical framework of facilities management based on available literature and principles. Chen et al. (2007) informed further adoption of the tool in studies that require integrated systematic literature review as a foundation for scope of scope and direction of research. Consequently, the current study adopted NSP as a tool for research focusing on themes associated with TM approaches for BIM pervasive major project delivery.

TM techniques facilitate a knowledge-driven systematic literature review. TM for BIM adoption in major projects supports decision making over the lifecycle of projects, and facilitate more effective management of projects (Chen and Bareka, 2022). The necessity of new research to enhance performance in major construction project delivery emphasises the essence of TM techniques. These views informed the adoption of an EBL approach to guide a consistent and reliable assessment in developing a study roadmap for performance enhancement in BIM pervasive major project delivery (Chen et al., 2021). Combining several methods in the research demonstrated effective identification of research themes and areas with potential gaps that guided the research roadmap for TM approach for BIM oriented performance enhancement in major construction projects.

A holistic systematic literature review enabled by the NSP tool provided justification the aim and objectives of the research. Also, the review highlighted fundamental research themes and where there was a need to establish a new TAPE. Following a need to derive a reliable set of BIMTAPE Framework, it was necessary to base it on a comprehensive systematic literature review and practice and ensure its suitability to clustered themes and work stages throughout the lifecycle of major construction projects. For more focused research, as advanced in the current thesis, the NSP, a TRIZ tool, was selected to achieve these needs. NSP was able to provide a qualitative identification and consequent justification of the BIMTAPE Framework and research task clusters.

The NSP approach described in the current thesis is illustrated in Figure 2.1 (see Chapter 2). The diagram is comprised of nine squares, or windows, used to derive the framework of this study. Fundamentally, the NSP looks horizontally from three perspectives of the subject problem: the “history, present, and future”. This is achieved through microcosmic and macroscopic review of available information and a system level review vertically. The illustration of the NSP in Figure 2.1 indicates the set-up of BIMTAPE Framework and how the goal was achieved through the establishment of potentials in BIM execution for major construction project delivery within the central window, gathering input from the remaining seven windows through evidence-based assessment.

3.8.1.2 Qualitative TM Data Collection and Ontological Modelling

The qualitative TM data collection process in this research focuses on the analysis of unstructured textual data using the BIMTAPE framework. The framework addresses the limitations of traditional BIM standards, such as IFC, by incorporating advanced TM and NLP techniques. These techniques enable the analysis of unstructured data, such as project reports, emails, contracts, and other text-based sources, which are often overlooked in conventional BIM workflows.

3.8.1.2.1 Text Collection Using the BIMTAPE Framework

The text collection process was conducted using the BIMTAPE framework, a structured approach designed to handle the complexity of unstructured or semi-structured textual data. Textual data for this study was gathered from six key sources, referred to as BIM Text Sets (TS):

- BIM Text Set 1 (TS1): Academic articles on BIM
- BIM Text Set 2 (TS2): Technical files from real-world case projects
- BIM Text Set 3 (TS3): BIM standards from BSI
- BIM Text Set 4 (TS4): Approved documents for building regulations
- BIM Text Set 5 (TS5): Corporate BIM policies
- BIM Text Set 6 (TS6): National BIM strategy and policy papers

These text sets were carefully selected to provide comprehensive coverage of BIM-related themes and to ensure the relevance of the qualitative data to the study's objectives. The purposive sampling technique was employed, targeting authoritative sources that align with the scope of the research (Bilal et al., 2016).

The data collection process involved a combination of web scraping techniques and manual extraction. Python and R-Studio were utilised to scrape relevant online repositories and collect key documents from academic, governmental, and industrial databases. The gathered documents represented a wide spectrum of BIM-related knowledge, encompassing technical standards, policy papers, and case studies, ensuring that the textual data encompassed various facets of BIM.

3.8.1.2.2 Ontological Modelling within the BIMTAPE Framework

The BIMTAPE framework incorporates ontological modelling to structure the qualitative TM process. The framework is illustrated as a RIBA plan matrix (see Chapter 4), which was designed to address the complexity of data science projects focusing on text analysis. Although the framework is founded on the CRISP-DM model, a key distinction is the complexity of the textual data being analysed. The CRISP-DM approach typically involves six stages, but the BIMTAPE framework follows a more streamlined process with five stages: from BIM text collection to BIM-oriented insights.

Each stage of the framework is designed with a specific aim, task, and expected outcome, ensuring a clear progression in the development of the TM model. The BIMTAPE framework utilises ontology to establish relationships between different concepts within the text, which enables a deeper understanding of how BIM-related processes are described in major construction projects. Ontological modelling helps to organise and structure the data, enhancing the ability to extract meaningful insights from unstructured text.

The BIMTAPE framework is beneficial in that it enables repetition and iteration within each phase, facilitating the refinement of the TM model as new data is collected and analysed. These iterative cycles ensure that the framework adapts to the unique requirements of each dataset and construction project, making it a flexible tool for text analysis in BIM.

3.8.1.2.3 Text Pre-processing for Qualitative Analysis

Before applying TM techniques, the collected data underwent a pre-processing phase to convert the unstructured text into a format suitable for analysis. The pre-processing steps included:

- Tokenisation: Dividing the text into smaller units, such as words or phrases.
- Stop-word removal: Eliminating common, irrelevant words (e.g., and, the).
- Stemming and Lemmatization: Reducing words to their base or root form.
- Term Frequency-Inverse Document Frequency (TF-IDF): A weighting technique to assess the importance of words within the corpus.

These pre-processing steps transformed the unstructured data into a structured form, enabling deeper analysis through the application of TM algorithms (Nayak et al., 2016). By organising the data in this way, the BIMTAPE framework ensures that the extracted information is both relevant and actionable for BIM-related projects.

3.8.1.2.4 Qualitative Insights from TM

The application of TM techniques within the BIMTAPE framework allowed for the identification of key themes and topics from the collected data. Techniques such as keyword extraction, entity recognition, and topic modelling were employed to uncover patterns within the text. These insights were crucial in understanding the broader themes within the BIM pervasive project delivery domain, such as risk management, collaboration, and efficiency.

The insights gained from the qualitative analysis provided a foundation for the next stage of the research, where quantitative techniques, such as Social Network Analysis (SNA), were applied to validate and further explore the relationships identified in the text. The integration of ontological modelling within the BIMTAPE framework ensured that these insights were not only based on raw data but also supported by a structured conceptual understanding of BIM processes in major projects.

3.8.1.2.5 Privacy and Data Reliability Considerations

Throughout the data collection and analysis process, strict protocols were followed to ensure the privacy and reliability of the data. Given the sensitivity of TM and analysis, especially when dealing with large datasets from both public and private sources, it was imperative to follow established data privacy guidelines. The research adhered to protocols that ensured the credibility and authenticity of the collected documents, verifying that the data was drawn from reliable, authoritative sources and not random or opinion-based articles.

The use of the FAIR principle (Wilkinson et al., 2016) further supported the data handling process, ensuring that all data and metadata were findable, accessible, interoperable, and reusable. This approach helped maintain the transparency and integrity of the research, ensuring that the findings could be replicated and validated by other scholars.

The qualitative data collection phase, supported by ontological modelling within the BIMTAPE framework, provided a robust foundation for the analysis of unstructured BIM-related texts. By employing a structured, iterative approach to TM, this framework facilitated the extraction of meaningful insights that are critical for advancing BIM in the context of major construction projects.

3.8.1.3 Quantitative Data Collection and Analysis

In parallel with the qualitative insights gained through TM, Social Network Analysis (SNA) was employed as the quantitative technique. SNA is used to visualize and quantify relationships between different entities within a network. In this research, it was used to map the relationships between terms and themes identified during the TM phase, providing a quantitative validation of the qualitative findings.

3.8.1.3.1 Data Collection for SNA

SNA was applied to the terms and themes extracted from the TM phase. These terms were treated as nodes in the network, and their co-occurrence in the same documents was represented by edges. This mapping allowed the identification of central themes and the relationships between key concepts. By employing SNA, the research was able to measure how frequently certain terms appeared together and how closely related they were within the overall BIM framework.

3.8.1.3.2 Quantitative Analysis through SNA

SNA provided several key metrics to quantitatively validate the relationships identified during the TM phase:

- Degree Centrality: To identify the most frequently occurring terms or concepts in the network.
- Betweenness Centrality: To evaluate how often certain terms served as intermediaries between others.
- Closeness Centrality: To measure the proximity of certain terms to all others in the network, indicating their influence.

These metrics quantitatively demonstrated the importance of concepts such as collaboration, efficiency, and risk management in BIM-related discussions, complementing the qualitative insights provided by TM. SNA allowed for a clear, data-driven understanding of the interconnections between different BIM themes and their relative importance (Pryke, 2012).

3.8.1.3.3 Integration of TM and SNA

The combination of TM and social network analysis provided a robust mixed-methods approach. While TM offered qualitative insights by identifying and categorizing themes from unstructured textual data, SNA validated these findings through quantitative measures. The integrated approach ensured that both the depth and breadth of the data were analyzed thoroughly, offering a comprehensive understanding of the role of BIM in major construction projects (Bilal et al., 2016).

3.9 Justifying the Single Case Study Approach

The use of a single case study in this research is methodologically justified by adhering to established guidelines in case study research. Yin (2018) posits that a single case can be appropriate when it involves a critical, unique, or revelatory case, or when it serves as an exemplar for testing or developing theory. In this research, the Crossrail project (BIM Text Set 2) represents a critical case due to its complexity, scale, and the wealth of technical documentation available from a completed BIM-enabled project. Given that Crossrail is one of the largest infrastructure projects in the UK, it provides a unique opportunity to apply TM techniques within Building Information Modelling (BIM) to investigate how unstructured data can enhance project management and decision-making processes.

Eisenhardt (1989) supports the use of case studies for theory building, particularly in exploratory contexts where detailed, in-depth analysis of a phenomenon can lead to novel insights. The experimental focus of this research is on integrating TM into BIM workflows, where the single case study of Crossrail provides a practical, real-world context to test and refine these methods. This approach aligns with the need for contextual depth in understanding complex, project-specific processes, which are better captured through case study methodology rather than large-scale quantitative studies (Flyvbjerg, 2006).

Furthermore, Siggelkow (2007) emphasises the value of a single case in illustrating a point or demonstrating how a new theoretical concept might operate in practice. The Crossrail project's extensive technical documentation provides a rich dataset for applying TM techniques, allowing for a granular exploration of how unstructured textual data can complement the structured data typically handled by BIM systems. The selection of a single case enables this research to focus deeply on understanding the intricacies of applying TM in a completed, real-world project, offering critical insights for future research and practice in the integration of TM and BIM.

Gerring (2004) also highlights that a single case study is valuable for intensive analysis when the goal is to gain a deep understanding of a particular phenomenon rather than generalising across cases. In the context of this research, the integration of TM into BIM processes is an emergent area, and the detailed study of a high-profile case like

Crossrail provides the necessary foundation to develop a methodological framework. This framework could then serve as the basis for future multi-case studies or broader applications of TM in BIM across different projects.

Lastly, Stake (1995) argues that single case studies are suitable for instrumental purposes, where the case is used to gain insight into an issue or to refine a theory. The Crossrail project is instrumental in demonstrating the value of TM in BIM processes, and the lessons learned from this single case are expected to inform future applications of TM in other large-scale construction projects. By focusing on a single case, this research can provide rich, contextualised insights that might be lost in a broader comparative study, while still contributing to theory development in the fields of TM and BIM.

COVID-19 Impact Justification: The challenges posed by the COVID-19 pandemic further justify the use of a single case study in this research. As outlined in the University of Strathclyde's Covid-19 Impact Statement (2021), PGRs were encouraged to review and adjust their research plans in response to the disruptions caused by the pandemic. This included considering changes to methods, research scope, and data collection strategies to mitigate the impact of COVID-19 restrictions. A full version of the Covid-19 Impact Statement can be found under declaration section of this thesis for further reference.

Before the pandemic, the research plan may have included the possibility of conducting fieldwork or exploring multiple case studies to gather primary data. However, the restrictions on travel, access to sites, and face-to-face interactions severely limited these options. In light of these constraints, the research adapted by focusing on the Crossrail case study, which offers a rich dataset of secondary data and technical documentation that was accessible without the need for in-person site visits or interviews.

The detailed and context-specific nature of the Crossrail case allowed the research to continue with a more focused approach. Although the dataset was reduced due to limited access to primary data during the pandemic, the available documentation from Crossrail provided sufficient depth for a comprehensive analysis. This shift in focus aligns with the guidance outlined in the Covid-19 Impact Statement, which encourages flexibility in research design and methodology to accommodate the challenges brought on by the pandemic.

In summary, the selection of a single case study of Crossrail is both methodologically sound and a necessary adjustment in response to COVID-19 restrictions. It allows for a deep, contextual exploration of how TM can be integrated with BIM, while also ensuring that the research remains feasible and robust despite the challenges of the pandemic. The decision to focus on Crossrail, with its wealth of technical documentation, helped mitigate disruptions caused by the pandemic and allowed the research to continue progressing towards meaningful insights in this emergent area of study.

3.10 Exploration of Research Limitations and Implications

The research methodology used in this study has a number of limitations, as explained below.

The systematic literature review was confined to a specific timeframe and a predefined set of keywords. As a result, the outcomes of the study might lack generalizability to different timeframes or alternative keyword sets. To overcome this limitation, future research could consider undertaking a more extensive and comprehensive systematic literature review.

The ontological approach to BIMTAPE Framework is in its developmental stages and has yet to gain widespread usage. This limitation implies that the study's findings might not be readily applicable to other projects employing the BIMTAPE Framework. To tackle this constraint, future research could focus on formulating and validating a more robust ontological approach to BIMTAPE Framework.

The TM experiments were limited to a small dataset and may not be readily applicable to other projects. Consequently, the study's outcomes might not be broadly generalizable to projects utilizing TM. To overcome this limitation, future research could undertake TM experiments on a more extensive dataset, enhancing the potential for broader applicability.

Despite these limitations, the research methodology used in this study provides a valuable foundation for future research on the use of TM for performance enhancement in BIM. Future research could address the limitations of this study by:

- Conducting a more comprehensive systematic literature review.
- Developing and testing a more robust ontological approach to BIMTAPE Framework.
- Conducting TM experiments on a larger sample of data.

The findings of this study suggest that TM can be a valuable tool for improving the performance of major projects. Future research aiming to leverage TM for enhancing performance in BIM pervasive major project delivery holds the potential to offer substantial advancements to the field.

3.11 Ethical Considerations for the Research

Ethical considerations are an integral part of any research, particularly when it involves the collection, analysis, and presentation of data. In this study, ethical concerns were thoroughly addressed to ensure the integrity of the research process and the protection of all stakeholders involved. The key ethical principles followed in this study are based on the guidelines provided by the British Educational Research Association (BERA) and the Economic and Social Research Council (ESRC), ensuring compliance with established ethical standards in academic research.

3.11.1 Data Confidentiality and Anonymity

The primary data sources in this study, such as academic articles, policy papers, and technical reports, were all publicly available documents. Nevertheless, ethical considerations regarding the treatment of data were upheld. TM was used to analyze unstructured textual data extracted from these sources, and measures were taken to ensure that any proprietary or sensitive information was treated with confidentiality. For instance, in cases where internal or technical documents from companies were included in the analysis, only publicly accessible information was utilized to avoid any breach of confidentiality. To maintain data integrity and privacy, all sources were anonymized in the reporting process, with no proprietary data linked to specific organizations or individuals.

3.11.2 Informed Consent and Permission

Although this study primarily relied on secondary data sources, the principle of informed consent remained relevant, particularly in accessing certain technical documents or reports from public domains. Where necessary, permissions were sought for the use of documents that were not freely available in the public domain, and appropriate attributions were made to honour intellectual property rights. The systematic literature review and TM procedures adhered to the ethical standards for literature-based research, ensuring proper citation and acknowledgment of all authors and original sources.

3.11.3 Accuracy and Integrity of Data

The methodology of the research, particularly the TM and SNA techniques, was rigorously applied to ensure the accuracy and integrity of the data analysis. This ethical consideration is critical in research that involves computational techniques and quantitative analysis, as any misuse of data or misrepresentation of results could lead to misleading conclusions. The researcher adhered to best practices for NLP and TM to minimize biases in data interpretation and to avoid manipulation of the data. All data analyses were conducted transparently, ensuring that the findings are replicable and verifiable by other researchers.

3.11.4 Ethical Use of Algorithms and AI Tools

The use of ML and AI tools, such as those employed for TM, raises ethical questions around the transparency and fairness of the algorithms used. In this study, ethical AI practices were followed to ensure that the algorithms deployed did not introduce bias or discrimination in the data processing and analysis stages. The TM models were trained on neutral, publicly available datasets to ensure fairness in the identification of patterns, themes, and relationships within the textual data. The algorithms were tested and validated for accuracy, with the findings cross-checked through SNA to ensure that they reflected the true relationships within the data, free from computational biases.

3.11.5 Data Storage and Protection

All data collected and analyzed during the course of the research were securely stored in compliance with data protection regulations such as the General Data Protection

Regulation (GDPR), which governs the handling of personal and sensitive data in the European Union. Although this study did not involve the collection of personal data, the protection of intellectual property and organizational data was prioritized. The data were encrypted and stored in secure databases, accessible only to the researcher. Additionally, all data files were anonymised, and unnecessary personal or sensitive details were excluded from the final dataset used for analysis.

3.11.6 Transparency and Ethical Reporting

Transparency in reporting the research findings was maintained throughout the study. The methodologies used, particularly SLR, TM, and SNA, were clearly outlined to ensure that readers could follow and replicate the research process. All findings, whether positive or negative, were reported in full, and there was no selective reporting of results to fit preconceived conclusions. Additionally, conflicts of interest were disclosed, and the researcher remained impartial and objective throughout the research process.

3.11.7 Academic Integrity and Avoidance of Plagiarism

In adhering to ethical academic standards, plagiarism was strictly avoided. All sources of information, from academic literature to industry reports, were duly cited using the Harvard referencing style, and ideas from other researchers were properly acknowledged. The researcher ensured that the original work was clearly distinguished from the work of others, and due credit was given for all contributions that informed the research findings. To maintain academic integrity, plagiarism detection software was employed to cross-check the thesis before submission.

The ethical considerations outlined above were rigorously applied at every stage of the research process. From the collection of unstructured textual data using TM to the quantitative validation using social network analysis, ethical best practices were adhered to, ensuring that the research was conducted with integrity, transparency, and respect for all parties involved. This commitment to ethical standards strengthens the credibility of the study’s findings and ensures that the research contributes positively to the growing body of knowledge on BIM and TM in the construction industry.

By following these ethical guidelines, this study not only provides reliable and valid results but also sets a foundation for future research that can build upon the methodologies and findings presented here.

3.12 Chapter Summary

This chapter detailed the comprehensive research methodology employed to address the objectives of this study, focusing on the integration of TM within BIM to enhance performance in major construction projects. Following the "research onion" framework by Saunders et al. (2016), the chapter outlined the philosophical foundation, the research approach, and the methods for data collection and analysis.

The study adopts pragmatism as the guiding research philosophy, allowing for the use of both qualitative and quantitative methods. The methodology employs a convergent parallel mixed-methods approach, combining TM for qualitative analysis of unstructured data and SNA for quantitative validation. This integration provides a robust framework for analyzing and interpreting the unstructured textual data gathered from BIM-related sources.

A SLR enabled by the NSP was conducted to establish a solid theoretical foundation for the study. The NSP allowed for a comprehensive review of existing literature, minimizing biases, and identifying gaps in the research related to BIM and TM integration. The SLR informed the empirical phase, where data collection and analysis techniques were further developed.

TM was applied as the primary qualitative tool, focusing on analyzing unstructured textual data such as academic articles, policy papers, standards, and technical reports. TM provided insights into recurring themes, patterns, and relationships within the data, while the use of NLP techniques facilitated the transformation of unstructured data into analysable formats. Six distinct text sets (BIM-TS1 to BIM-TS6) were selected to ensure comprehensive coverage of BIM-related topics.

For quantitative validation, SNA was employed to map the relationships between key terms and themes identified through TM. SNA allowed for the visualisation of connections between concepts and measured their centrality, betweenness, and closeness in the network. This provided a data-driven understanding of the interrelations within BIM-related discussions.

Ethical considerations, particularly in handling unstructured textual data and ensuring data privacy, were also addressed, ensuring that the research adhered to established guidelines and protected the confidentiality of the sources.

In conclusion, the mixed-methods approach, integrating TM and SNA, enabled a thorough investigation into how BIM and TM can enhance performance in major construction projects. By triangulating qualitative and quantitative data, the research methodology ensures the credibility, reliability, and validity of the findings, contributing to advancements in BIM practices. The methodologies and techniques used in this study provide a framework for future research on integrating digital technologies into construction project management. Chapter 4 lays the groundwork for the subsequent empirical investigation, setting the stage for the analysis and evaluation of textual data using the BIMTAPE Framework.

Chapter 4
BIMTAPE Framework and Experiments

4.1 Introduction

This chapter introduces the BIMTAPE Framework, a tool developed within this thesis to investigate the integration of TM with BIM for performance enhancement in major construction project delivery. While BIMTAPE is not a research methodology, it serves as an analytical framework that addresses the methodological and knowledge gaps identified in BIM research, particularly in dealing with unstructured textual data. The chapter provides an overview of each stage of the framework, detailing its methodologies, tools, and techniques. The BIMTAPE Framework bridges the gap between TM and BIM, facilitating more informed decision-making and improved project outcomes. It establishes a connection with previous chapters and sets the stage for applying the framework to analyse unstructured textual data from BIM-enabled major construction projects.

4.2 Text Analytical Challenges

TM encompasses the procedure of examining extensive quantities of unstructured text data and extracting useful insights and information. It has various applications in different industries, including entertainment, where it can be used to analyse user feedback and improve content quality (Ittoo et al., 2016). Companies like Netflix have already adopted TM techniques and have seen benefits from it (Amatriain and Basilico, 2012; Harris, 2014). The utilization of TM is anticipated to rise in the coming years (Harris, 2014), as organizations pursue to harness the abundant information contained within unstructured text data.

This research focuses on providing a framework for TM that can standardize the fundamental processes involved in the analysis of unstructured text data. The process of TM can present challenges due to the unstructured character of the data and the intricacy in deriving substantial insights from it (Harpaz et al., 2014). However, the methodology proposed in this research aims to overcome these challenges and provide a systematic approach for TM that holds the potential to yield outcomes that are both more effective and efficient. By following a standardized framework, companies can leverage TM to gain a competitive advantage and extract valuable insights from their data.

TM has its own set of challenges, mainly due to the nature of text data, which is often unstructured, coming from a variety of sources such as social media, voice recordings, etc. (Hassani et al., 2020). These sources store data in different formats, posing a challenge for data scientists aiming to merge and analyse them (Hassani et al., 2020). As a result, this necessitates several pre-processing steps to structure the data in a way that enables efficient analysis (Juneja et al., 2019). These challenges make TM a complex process, but despite these difficulties, TM has potential to support performance enhancement in various industries, including construction, by providing new insights and analysis from large amounts of text data (Wakefield and Bean, 2004).

Additionally, the use of slang, idiomatic expressions, and abbreviations can also add to the complexity of analysing text data. Furthermore, text data often contains irrelevant or redundant information, as well as inconsistencies and errors, which must be cleaned and processed before analysis can be performed (Sun et al., 2018). Therefore, it is crucial to apply the right methodology to tackle these challenges, such as the normalization of text, the removal of stop words, stemming, and lemmatization to make the text data amenable to analysis.

In many multinational companies, the use of English as a common language for business communication is widespread, especially at the higher management level. However, at the operational level, local languages are often used for communication as it can be more efficient and effective in conveying information to employees who may not be fluent in English. It also helps to maintain cultural diversity and create a more inclusive work environment (Ittoo et al., 2016).

Dealing with human-generated data in a natural language format can be highly challenging for automation processes. Natural language is complex and diverse, and understanding it requires a deep understanding of human language, including context, grammar, semantics, and cultural nuances. Additionally, the informality and sparsity of the data can make it difficult for machine algorithms to extract meaningful information (Rajput, 2020). The use of casual words and ill-formed grammatical constructs can also make it challenging for algorithms to accurately process the data

(Ittoo et al., 2016). This is a major hurdle in industries such as customer service, sentiment analysis, and information retrieval, where large amounts of unstructured text data are generated every day.

To overcome language-processing challenges, Natural Language Processing (NLP) uses advanced techniques to help machines better interpret human language (Ittoo et al., 2016). Named Entity Recognition (NER) identifies and classifies specific elements in text, such as names, dates, and locations, allowing systems to extract meaningful entities from unstructured data (Weischedel et al., 2013). Part-of-Speech (POS) tagging labels words in a sentence according to their grammatical roles like nouns, verbs, or adjectives, providing structural insights that assist in parsing sentence meaning (Jurafsky and Martin, 2019). Sentiment Analysis assesses the emotional tone within text, determining whether the expressed opinion is positive, negative, or neutral, which is particularly useful for gauging user attitudes or public sentiment on a topic (Liu, 2012).

Furthermore, in the context of text analytics systems in BIM pervasive major project delivery, domain knowledge expertise is critical for accurately capturing the specific industry keywords used. BIM projects, construction or infrastructure initiatives that use BIM technology to create and manage a digital representation of the building or infrastructure, generate large amounts of text data, including project specifications, contracts, drawings, and reports (Eastman et al., 2011). This text data is an important source of information for decision-making and project management.

However, to effectively analyse this text data, the text analytics system must be capable of understanding the construction industry's specific language and terminology, which includes diverse and often context-specific terms related to project phases, materials, methods, and regulations. Rather than depending on structured data standards like IFC or bSDD, this system processes unstructured data directly through TM and NLP techniques, enabling it to capture meaningful industry-specific insights from textual sources. This requires domain knowledge expertise and a robust NLP engine to accurately identify and extract relevant information from the text data (Ittoo et al., 2016). Having such a system in place can help organizations to improve project delivery, reduce risks, and make data-driven decisions in a more cost- and time-efficient way. Nowadays, companies are increasingly investing in new technologies that can deliver value in real-time or near-real-time to support fast-paced decision-making (Zohuri et al., 2022). This is especially true for text data, where organizations need to process vast amounts of information in a timely manner to make informed decisions.

Speed is a key requirement for many organizations in the design of analytics solutions that address the challenges of Big Data complexity (Naeem et al., 2022). This has led to the development of advanced NLP techniques and algorithms that can process text data in real-time or near-real-time, providing organizations with the ability to make data-driven decisions quickly and efficiently. Additionally, advancements in cloud computing, distributed systems, and parallel processing have also made it possible to scale text analytics solutions to meet the needs of organizations with large amounts of text data (Ittoo et al., 2016). The rapid data generation rate poses a significant challenge that text analytics systems must address.

Social media networks, such as Twitter, are a prime example of this challenge, as they produce massive amounts of text data in real-time. As per Twitter Usage Statistics (2022), on average, 6,000 tweets are generated every second on Twitter, creating a vast and constantly-changing stream of information. This requires text analytics systems to be highly scalable, efficient, and capable of processing large amounts of data in real-time.

The analysis of text data presents unique challenges that go beyond standard data processing. In research, surveys like the one by Shen et al. (2022) have explored tools for extracting value from natural language data. For instance, Tan (1999) compiled a list of TM products and applications for refining text and distilling knowledge. In the context of BIM, adopting new processes and tools is crucial for supporting large teams and enhancing project effectiveness. This requires a solid understanding of NLP techniques and their integration into BIM workflows to extract insights from text data.

Additionally, the development of BIM-specific text analytics systems that are designed to meet the unique needs of the construction industry can also help to improve project delivery and decision-making (Saltz, 2015). The step-by-step process of text analytics often assumes that the results will be error-free, but in reality, this is not always the case. While data scientists acknowledge the significance of producing high-quality outcomes, there has been relatively less emphasis on establishing a well-defined process to ensure such quality results. This stands in contrast to the thorough methodologies commonly employed in software development.

Therefore, the proposed BIMTAPE Framework aims to address the challenges associated with text analytics in BIM-oriented major construction project delivery. The framework considers the integration of ML methodologies and text analytics with Big Data mining to enhance the performance of BIM-oriented major construction project delivery. The main contribution of this thesis is the proposal of a new technical solution in the context of text analytics and its integration with BIM to improve performance and decision-making in major construction project delivery. The BIMTAPE Framework is designed to provide a structured, systematic approach to text analytics that considers the unique needs and requirements of the BIM industry and is intended to help organizations to better leverage the potential of text data to support project delivery and decision-making.

4.3 Development of BIMTAPE Framework

The BIMTAPE Framework serves as an analytical tool developed to address the challenges of text analysis in BIM-oriented major construction projects. Although it is not a research methodology, it is designed to complement existing methodologies by offering a structured and systematic approach to text analysis in BIM. The framework consists of a set of steps that guide the user through the process of identifying the purpose of the text, understanding the context in which it was produced, and evaluating the quality and relevance of the information it contains.

By following the steps of the BIMTAPE Framework, organizations can more effectively extract valuable information from BIM-related texts and use it to support decision-making and project delivery. The framework is made to be flexible and adjustable according to the specific needs of different organizations and projects and can be used to analyse a wide range of BIM-related texts, including technical documents, specifications, contracts, and standards. The BIMTAPE Framework was developed to address the challenges of text analysis in the context of BIM and to enhance performance in BIM-related construction projects.

It relies on the framework of the Cross Industry Standard Process for Data Mining (CRISP-DM) methodology, which provides a structured and systematic approach for conducting and organizing data mining projects (Chen et al., 1996). CRISP-DM was developed by a consortium of five companies, as described in Chapman et al. (2000), including Daimler AG, Teradata, Integral Solutions Ltd (ISL), OHRA (an insurance company), and NCR Corporation. It has become a widely adopted standard for data mining projects and provides a structured approach for conducting data mining projects from start to finish (Piatetsky, 2022).

Systematic documentation along with effective project management play an important role in the success of KDD projects (Skarpathiotaki and Psannis, 2022). A meticulously organized project management process aids in guaranteeing timely delivery, adherence to budget constraints, and achievement of the necessary quality benchmarks for the project. Systematic documentation helps to ensure that the results of the project are well understood and can be easily communicated to stakeholders (Tan, 1999). Despite numerous attempts to introduce various approaches for overseeing data mining projects, the lack of a proper methodology for project development is one of the common pitfalls in data mining projects (Moss and Are, 2003).

Although there is a well-defined methodology for conducting KDD projects, numerous real-world projects still tend to be approached in a disorganized and improvised manner (Becker and Ghedini, 2005; Nadali et al., 2011). This is likely due to the complex and ever-evolving nature of the field, as well as the diverse range of techniques and tools available for conducting KDD. Nevertheless, it is important for practitioners to follow a structured methodology to ensure the validity and reliability of their results. The six phases of KDD as adumbrated by Chapman et al. (2000) are as follows:

- **Phase 1: Business Understanding:** The goal in this phase is to understand the business problem that the data can solve. This includes defining the objectives and goals of the project, as well as identifying the stakeholders who will be impacted by the results.

- **Phase 2: Data Understanding:** The goal in this phase is to get a comprehensive understanding of the data that will be used for the project. This includes exploring the data to identify patterns, outliers, and other characteristics.
- **Phase 3: Data Preparation:** The goal in this phase is to clean and prepare the data for analysis. This includes dealing with missing values, correcting errors, and transforming the data into a format that can be used for analysis.
- **Phase 4: Modelling:** In this phase, the goal is to use the data to build models that will help solve the business problem. This includes selecting the appropriate algorithms and tools, as well as testing and validating the models.
- **Phase 5: Evaluation:** In this phase, the goal is to evaluate the effectiveness of the models built in the previous phase. This includes comparing the results of the models to the objectives and goals defined in the business understanding phase.
- **Phase 6: Deployment:** In this phase, the goal is to deploy the solution to the stakeholders. This includes communicating the results, as well as implementing the solution in a real-world setting.

The CRISP-DM methodology, which is a widely-used process for conducting KDD projects, is conceptualized as an iterative cycle. The CRISP-DM methodology emphasizes the iterative and flexible nature of the KDD process and allows practitioners to make adjustments and refinements as necessary to achieve the best possible results.

CRISP-DM is a tool-agnostic procedure, meaning that it does not dictate the use of specific tools or technologies. This makes it a flexible and adaptable framework for conducting KDD projects. SPSS, which was later acquired by IBM, built upon the CRISP-DM methodology to create the Analytics Solutions Unified Method for Data Mining/Predictive Analytics (ASUM-DM) (Chapman et al., 2000). ASUM-DM extends the CRISP-DM process by incorporating additional steps and best practices for conducting KDD projects using IBM’s analytical tools and technologies. However, at its core, ASUM-DM is based on the same iterative and flexible approach as CRISP-DM.

ASUM-DM has five phases: Analyse Design, Configure and Build, Deploy, and Operate and Optimize. These phases provide a structured approach to conducting KDD projects using IBM's analytical tools and technologies (Chapman et al., 2000). Other companies, such as SAS, have their own methodologies for conducting KDD projects. SAS has the SEMMA (Sample, Explore, Modify, Model, and Assess) methodology, which is similar to the CRISP-DM and ASUM-DM methodologies in that it provides a structured and iterative approach to conducting KDD projects (Chapman et al., 2000).

CRISP-DM and ASUM-DM are process-oriented methodologies for conducting KDD projects and provide a structured approach to the overall process of finding valuable insights from data (Schwade, 2021). There are also other approaches that are more programming-oriented, and one such approach is the use of R technology for data science. The use of R is intricately connected with the suite of tools created by RStudio, which are widely used for text categorization and analysis (Giorgi et al., 2022).

The R-based approach to KDD emphasizes the iterative aspect of data exploration, and involves repetitive steps such as Visualize, Model, and Transform. This approach is tempting due to its simplicity, but it is important to remember that other factors such as validation, insights gained, and others, also need to be considered in parallel with the programming effort (Li et al., 2018).

For practical purposes, the choice of methodology for conducting KDD projects will depend on the specific needs of the project, the skills and experience of the practitioners, and the tools and technologies available (Chapman et al., 2000). Both process-oriented methodologies like CRISP-DM and ASUM-DM, and programming-oriented approaches like the R-based approach, have their own strengths and weaknesses, and the right approach will depend on the specific requirements of the project.

There is no universally accepted methodology for TM, given that data mining techniques are typically tailored for extracting information from structured databases, whereas TM techniques are specialized for extracting information from unstructured textual data (Skarpathiotaki and Psannis, 2022). NLP is an important tool for enhancing information extraction procedures in TM (Rajman and Besançon, 1998).

Therefore, this study developed the customized BIMTAPE Framework for performance enhancement in BIM pervasive major project delivery, which is inspired by the CRISP-DM framework. This framework provides a generic structure for TM projects, and specific projects may require some changes and adaptations to the framework. The focus of the study is on methodology-oriented development of the BIMTAPE Framework.

The main goal of the proposed BIMTAPE Framework is to improve the performance of BIM pervasive major project delivery. The framework provides a systematic and structured approach to analysing BIM-related texts, such as standards, contracts, technical documents, and specifications, in order to extract valuable insights and enhance decision-making in construction projects.

The BIMTAPE Framework aims to address the identified research gap by leveraging TM techniques and applying them specifically to BIM-related texts. By utilising the framework, project stakeholders can gain a deeper understanding of the information contained within these texts, including its purpose, context, relevance, and quality. This enhanced understanding enables more informed decision-making, effective communication, and improved collaboration among stakeholders.

The framework also facilitates the identification of key concepts, patterns, and relationships within BIM-related texts through the application of TM methodologies. By analysing and interpreting the data extracted from these texts, the framework helps to identify potential problems, optimize construction processes, and enhance overall project outcomes.

Ultimately, the goal of the BIMTAPE Framework is to provide project stakeholders with a valuable tool for analysing and utilizing BIM-related texts to support performance enhancement in BIM pervasive major project delivery. By integrating TM techniques with BIM-specific knowledge and expertise, the framework aims to bridge the gap between theoretical insights and practical application, enabling stakeholders to harness the full potential of BIM technology and improve project performance.

The proposed BIM text analytics framework is illustrated in Table 4.1 using the RIBA plan matrix, which is a well-known framework used in architecture and construction projects (RIBA, 2015). This framework is specifically designed to address data science projects cantered around text analysis and is rooted in the CRISP-DM methodology. The central objective of the introduced BIMTAPE Framework is to present five conceptual pathways for advancing a TM model. The primary distinction between this framework and the conventional CRISP-DM method lies in the composition and intricacy of text data. This complexity necessitates an alternative approach to the process of data analysis. The RIBA plan matrix is a useful tool for organizing and structuring the steps involved in a data science project and helps to ensure that the project is carried out in a systematic and organized manner (RIBA, 2015). By following this framework, practitioners can make the most of the information contained in the text data and extract valuable insights that can be used to improve their processes and decision-making.

The BIMTAPE Framework is a customized approach to the CRISP-DM methodology that has been specifically designed for text analysis in the field of BIM. The BIMTAPE Framework consists of five stages, starting from BIM text collection and ending with BIM-driven insights gained. Each stage of the BIMTAPE Framework has its own aim, tasks, and outcome. The framework is used to trace the progress of the model development, and as with other methodologies, it is based on a series of iterations. Each phase may require one or more iterations, and the framework is designed to be flexible enough to accommodate different requirements for different projects. The five stages of the BIMTAPE Framework are:

- Stage 1: BIM text collection
- Stage 2: Pre-study BIM business understanding
- Stage 3: BIM TM data collection and understanding
- Stage 4: Validation of BIM TM approach
- Stage 5: Performance-boosting BIM insights.

By following the BIMTAPE Framework, practitioners can ensure that they are collecting the right data, understanding it properly, validating their models, and extracting valuable insights that can help improve their processes and decision-making in BIM-related major projects.

4.3.1 BIMTAPE Framework Conceptual Development

The development of the BIMTAPE Framework draws on theoretical concepts from TM, NLP and BIM. However, these fields have developed methodologies for unstructured data associated with construction projects, which have led to better collaboration, decreased errors, and enhanced performance (Guo et al., 2022; Zhou et al., 2019). However, the BIMTAPE framework embraces the void in information handling of unstructured textual data which involves project reports, emails and contracts (Hosseini et al., 2017).

TM and NLP Foundations: TM and NLP help get value from the huge tonnage of unstructured data. Here, theoretical concepts from data mining and knowledge extraction models are applied to help identify patterns in project communications and documents that may aid in decision making (Aggarwal and Zhai, 2012). In particular, the use of the Cross-Industry Standard Process for Data Mining (CRISP-DM) to define a structured procedure for data analysis (Mazairac and Beetz, 2013) has a powerful influence on the design of the BIMTAPE framework. Directly applied in BIMTAPE, this standard has stages which include data preparation, modeling, and evaluation.

The framework is not specific to structured data model IFC, however TM and NLP complement BIM bringing additional insights. While BIM focuses on the structured data in 3D models and schedules, TM offers the possibility of analyzing project related texts, which many BIM standards neglect. BIMTAPE improves project performance beyond the point where purely structured standards can achieve (Karan and Irizarry, 2015; Tang et al., 2019), by extracting and analyzing unstructured data.

4.3.2 Rationale for Excluding IFC and bSDD

The Industry Foundation Classes (IFC) and building SMART Data Dictionary (bSDD) are widely recognised standards in BIM, According to Kumar (2024). However, they focus on structured data, which differs significantly from the unstructured textual data examined in this research. The exclusion of IFC and bSDD in the BIMTAPE Framework is justified based on their limitations in dealing with unstructured data.

4.3.2.1 Nature of IFC and bSDD: Structured Data Models

IFC and the bSDD are widely recognised as foundational standards in BIM (buildingSMART, 2020; Eastman et al., 2011; Kumar, 2024). These standards support structured data exchange and collaboration, playing a critical role in ensuring consistency and standardisation in BIM workflows (Pauwels and Zhang, 2015; Succar, 2009). However, their primary focus is on structured data, which differs significantly from the unstructured textual data that forms the core of this research. IFC is a schema designed to store structured data related to building components, geometrical properties, and schedules, making it ideal for organising static and hierarchical information (Karan and Irizarry, 2015). Similarly, bSDD provides a standardised vocabulary for BIM data, facilitating interoperability between BIM platforms.

4.3.2.2 Emphasis on Unstructured Data

Unstructured data, such as project reports, communications, and meeting minutes, constitutes a substantial portion, often over 80%, of project-related information in the AEC industry (Guo et al., 2022). This data is highly dynamic and context-dependent, making it difficult to manage using traditional structured data models. TM and NLP techniques are well-suited for analysing unstructured data, enabling the extraction of patterns, trends, and actionable information that structured data models like IFC cannot effectively handle.

4.3.2.3 Theoretical Justification: Gaps in IFC and bSDD

The exclusion of IFC and bSDD is further justified by their limitations in handling unstructured data. While IFC is an effective tool for structured data representation, its static and hierarchical nature makes it ill-suited for managing the evolving and non-parametric nature of unstructured data (Karan and Irizarry, 2015). Recent research by Sobhkhiz and El-Diraby (2023) emphasises that linking structured IFC data to unstructured textual data is insufficient due to the rigid relational schemas that fail to account for the complex, dynamic relationships within unstructured textual data.

Additionally, approaches for integrating unstructured data into IFC models often rely on pre-defined taxonomies or relational mappings. These approaches impose constraints on the analysis of unstructured data, as they require extensive manual coding or adherence to standards that may not reflect real-time contextual shifts in project documentation. Thus, excluding IFC in the current research allows the framework to focus on extracting relationships and conceptual insights that are critical for understanding unstructured data but cannot be captured by rigid structured models.

4.3.2.4 Future Path: Hybrid Model for Structured and Unstructured Data

While the exclusion of IFC and bSDD is justified in this research, the study acknowledges their importance in BIM workflows. A future enhancement to this framework could involve developing a hybrid model that integrates both structured and unstructured data, leveraging the strengths of each.

4.3.2.5 Semantic Networks and Conceptual Distances

A hybrid model could be implemented through the use of semantic networks and conceptual linking mechanisms. Semantic networks could be constructed from unstructured text to represent relationships between project concepts, which can then be mapped to IFC classes by calculating conceptual distances. This approach would allow the system to automatically link unstructured text data to relevant IFC objects, thereby bridging the gap between structured BIM data and the unstructured project documentation.

4.3.2.6 Taxonomy-Based Integration

An additional layer of integration could involve the use of taxonomies or ontologies to support the interoperability between structured and unstructured data. TM and NLP techniques could be used to annotate unstructured text with relevant IFC codes, creating a taxonomy-based relational mapping that facilitates smoother transitions between data types.

4.3.2.7 Mitigating the Risk of Exclusion

While the decision to exclude IFC and bSDD may appear risky given their widespread adoption in BIM workflows, this risk is mitigated by the clear academic and methodological foundation of the research. As shown in recent studies, IFC's inability to handle unstructured data limits its usefulness in certain analytical contexts (Tang et al., 2019). By focusing on unstructured data, this study addresses a critical gap in current BIM practices, offering a unique perspective that complements traditional IFC-based approaches (Sobhkhiz and El-Diraby, 2023). Furthermore, the future development of a hybrid model ensures that this research remains relevant to industry practices while pushing the boundaries of what BIM can achieve by integrating both structured and unstructured data types.

Table 4.1: BIMTAPE Framework

Stages and Aims	Stage Tasks	Stage Outcomes
1. Text collection		
To extract texts from organizational databases and other sources of information.	TA1.1: A chain of evidence such as the 6 types of documentation to be integrated as sources of documentation. TA1.2: Extraction of texts.	Textual data that can be used to gain insight about the processes in the organization.
2. BIM Use Case and Business Understanding		
To gain understanding about the nature of the organization's processes, as well as its stakeholders and their needs.	TA2.1: Identification of key stakeholders TA2.2: Study of organizational processes TA2.3: Understanding stakeholder needs.	An overview of organizational processes, stakeholders, and their needs.
3. BIM TM data collection and understanding		
To gain an understanding of the data and its properties, including identifying the structure and patterns in the data.	T1: Data pre-processing T2: Data exploration T3: Identification of patterns and structures in the data	The acquisition of unstructured text that can be analysed to understand the state of the organization.
4. Validation of BIM TM approach		
To evaluate the models generated in previous stages and determine their validity.	T1: Model evaluation T2: Validation of results T3: Refinement of models	Validated models that can be used for further analysis.
5. Performance-boosting BIM insights		
To extract insights from the data and use them to make informed decisions.	T1: Documentation of T1: Interpretation of results T2: Identification of key insights T3: Generation of recommendations	Key insights and recommendations that can be used to drive decision making.

4.3.3 BIM Text Collection

BIM text collection stage has aim, tasks, and expected outcomes. The main aim of this stage is to establish BIM Text Sets (TSs) for data collection. For this research to be more comprehensive, BIM TS collection in connection to BIM practices in major projects consists of six BIM TSs, and they are:

- BIM TS 1 Academic papers
- BIM TS 2 Major project learning legacy files
- BIM TS 3 Regulation documentation
- BIM TS 4 British standards
- BIM TS 5 National BIM policy papers
- BIM TS 6 Enterprise BIM policy papers

The next step after defining the search boundaries is to collect the text data from each of the six BIM TSs. This can be done through various methods such as manual collection, web scraping, or utilizing existing databases. The collected text data should then be stored in a format that is appropriate for further analysis. It is important to ensure the quality and validity of the collected data by using methods such as data cleaning and data pre-processing. The collected text from each TS will undergo various stages of analysis, ultimately contributing to a holistic understanding of BIM-related insights, from compliance and regulatory alignment to organisational practices and theoretical frameworks.

The results of each stage will provide a deeper understanding of BIM processes in major projects and ultimately support enhanced performance in these projects. The data collected from each TS will also be used to verify and validate the findings of previous studies in the field. The use of multiple BIM TSs helps to increase the validity and reliability of the research findings. The outcome of this stage is a set of text data that is ready to be analysed in the next stage of the BIMTAPE Framework.

In this stage, TM is used to derive valuable information from the collected data in an automated and systematic manner. After obtaining data from the above-mentioned sources, a set of filtration process on all BIM TSs were carried out to select the most relevant papers related study. Stated in detail for clarity, bibliometric and trend analysis were carried out based on BIM TS 1 Academic papers only (retrieved data from Scopus and WoS). The main reason for considering BIM TS 1 for both bibliometric and trend analysis is because these databases provide high-quality data (e.g., peer-reviewed studies from respected academic journals), and permit the option to refine search queries and perform search in the search string. This also allows access to the world's leading research analytical information and scientific citation platforms (Chavarro et al., 2018; Li et al., 2018).

Using BIM TS 1 for both bibliometric and trend analysis allows for a more focused analysis on academic papers, which are likely to provide valuable insights into the current state of BIM research and its evolution over time. This will help in understanding the current trends and advancements in the field of BIM and text analysis. The results from bibliometric and trend analysis provides a better understanding of the research landscape and identify key authors, institutions, and journals in the field of BIM.

This research study is focused on text analysis of research articles. The data collected from Scopus and WoS databases needs pre-processing before text analysis can be performed. The focus of this study is on the full texts of research articles and the pre-processing techniques used include bibliometric analysis and trend analysis to identify general trends in BIM use for major construction project delivery. The following sections describe the focus of the research study, and compare techniques used for text analysis to determine the most applicable tools for BIM-related text data. The aim of this stage is to analyse the current state of BIM practices and their impact on the performance of major construction project delivery, as well as to identify potential future directions for more efficient and effective application of BIM methodologies and tools.

4.3.4 BIM Use Case and Business Understanding

Since this study undertook a critical systematic literature review using a systematic approach using NSP, therefore the BIM business understanding stage was developed to capture the use of BIM in practice. The goal of BIM business understanding stage was developed to capture the use of BIM in practice and gain a deeper understanding of real-time BIM scenarios. At this stage, the researcher used web scraping to examine

BIM TS 2 Major project learning legacy files and employed a Python package called "BoilerPy3" for natural language analysis. The goal of this stage was to develop a text analytics model that addresses the knowledge and information needs of the organization and provides relevant information for improving their understanding of construction functions and processes. The aspect of business understanding focused on ensuring that the users of the text receive the necessary information to improve their insights and understanding of the organizational functions and processes.

This passage explains the source of information used in the experiment and the methods used to gather it. The experiment used data from the "crossrail.co.uk" webpage as a case-based example in the context of major construction project delivery. The data was extracted using web scraping with Python and a variety of techniques and tasks were followed, as described in Table 4.2.

The expected outcome of this stage is to identify the key factors that influence the effectiveness and efficiency of BIM adoption in major construction projects, as explored through a detailed analysis of a single, real-time case project. The use of a single case study is justified and thoroughly outlined in the research methodology, drawing on established case study principles such as those by Yin (2018), which support single-case studies in situations that require in-depth, context-rich analysis. This approach enables a comprehensive examination of BIM's real-time application within a representative, complex project setting, offering insights that are both practically grounded and academically rigorous. Furthermore, the choice of a single case is reinforced by practical considerations, including COVID-19 restrictions and time limitations, which impacted the scope of data collection during the study period.

Table 4.2: Stage 2 of the BIMTAPE Framework for the BIM text extraction

	Stage 2: BIM business understanding
Aim	To identify business processes and activities affected by the information modelling.
Tasks	2.1: Identify organizational functions and processes. 2.2: Identify tools and resources necessary to understand organizational functions based on the data collected. 2.3: Develop a plan for text analytics according to organizational processes, needs and functions.
Outcomes	2.4 Improved knowledge and understanding of organizational needs, information, and data.

4.3.5 BIM TM Data Collection and Understanding

This section describes the "BIM Data Modelling (TM)" stage of the research. The main purpose of this stage is to build different TM models from the obtained BIM TS 3-BIM TS 6 textual data and start processing. The tasks (see Table 4.3) involved in this stage are:

- Model selection: There are many different procedures for model construction, and new models can be created by combining other models.
- Parameters tuning: Most model construction procedures have parameters that can be adjusted, and different strategies may be used to find the best set of parameters.

Formatted and saved text data was obtained from various sources (BIM TS 3 Regulation documentation, BIM TS 4 British standards, BIM TS 5 National BIM policy papers, and BIM TS 6 Enterprise BIM policy papers) (see Table 4.4). The data was formatted into data frames and saved in .pdf and .txt files for faster access during computational analysis. The pre-processing of the data was done using the RStudio NLTK (Natural Language Toolkit) library and Python. The pre-processing steps applied to the raw text data were:

- Step 1: Tokenization: The raw text was divided into smaller units called tokens.

- Step 2: Removal of punctuations and stop words: Punctuations and stop words were removed from the text.
- Step 3: Conversion to lower case: All words were converted to lowercase.
- Step 4: Lemmatization: The process of reducing words to their base or root form was performed for frequency counting.

Each experiment for BIM TS was undertaken separately, and the results of each experiment are presented. The modelling part is divided into two stages, including

- Exploratory Analysis: The word frequencies and co-occurrence were analysed to find the trend of occurrence for different BIM adoption issues related to performance enhancement in major construction project delivery. Visualizations such as bar charts, trending charts, and word clouds were used to represent the results.
- ML Techniques: Both supervised and unsupervised classification models were used to classify the text data. The classification models were used to categorize the text data into different classes or groups (Alloghani et al., 2020).

To evaluate the supervised classification models, the BIM TS 1 academic papers was used as the classification tags. A variety of techniques were used, including (1) Information Retrieval (IR) (2) keyword co-occurrence analyses, and (3) Document-Term Matrix (DTM). Text records must be parsed concerning decoding documents (analysed into logical syntactic components). The learning machine approach must help the researchers address the slits of in-service knowledge, describe new development themes, and reveal the hidden structures. Moreover, it gives out an assessment of the area researched and a holistic picture. The expected outcomes for this stage will be a set of different TM models and techniques for BIM pervasive major project delivery.

Table 4.3: The BIMTAPE Framework procedure stage

	Stage 3: Data understanding (TM)
Aim	To prepare the textual data and begin processing
Tasks	3.1 Model selection 3.2 Parameters tuning
Techniques	Information Retrieval (IR) Keyword co-occurrence analyses Document-Term Matrix (DTM)
Outcomes	TM models and technique for BIM pervasive major project delivery

Table 4.4: The data sources that will be used during TM stage

Source	**Type of Data**
BIM TS 3 Regulation docs	Regulatory documents
BIM TS 4 British standards	Standards and guidelines
BIM TS 5 National policy	Policy papers
BIM TS 6 Enterprise policy	Organizational policy papers

4.3.6 Validation of BIM TM Approach

The aim of the BIM model validation stage is to assess the effectiveness of the TM model in achieving the objectives of TM and analysis. The tasks involved include setting up the model, conducting an analysis of sample texts, and evaluating the model's ability to categorize texts, generate meaning, and provide new insights (see Table 4.4). The evaluation process involves comparing the TM performance through different SNA and manually verifying each word to determine if the TM model has accurately mapped words to specific topics/classes. The goal of the stage is to develop a TM approach for enhancing the performance of major construction project delivery in BIM. The process between BIM TS and the BIM model validation is iterative, seeking to improve results.

Table 4.5: The RIBA matrix for the new insights gained from the BIM model

	Stage 4: Validation
Aim	To use the BIM model to undertake text analysis and validation.
Tasks	4.1: Launch the model. 4.2: Enter the text in the model and conduct analysis. 4.3: Evaluate the effectiveness of the model in text analysis.
Outcomes	4.4 The model will be validated by its success in accurately categorising data, extracting relevant information, and generating insights that support enhanced project performance.

4.3.7 Performance-boosting BIM Insights

The final stage of the BIMTAPE Framework involves gaining insights from the model's results to support performance enhancement in BIM-pervasive major project delivery. This stage provides an opportunity to improve organizational functions such as building processes, communication, and resource management through the use of TM applications and visualization techniques. The results can be exported into a tabular format for further analysis by stakeholders and researchers, and various visualization techniques such as live maps, histograms, box-whisker plots, and scatter line plots can be used to present the insights.

4.4 Experiment 1: BIM TS 1 Academic Articles

4.4.1 Body of Text

4.4.1.1 'Data Collection

Due to the worldwide scope of BIM and the diverse range of relevant keywords, a user-friendly and accessible approach was necessary to merge datasets from Scopus and Web of Science (WoS) search results. This method was vital for effectively conducting bibliometric analysis techniques (Caputo and Kargina, 2021). The method employed was devised by Echchakoui (2020) and draws upon the method established by Caputo and Kargina (2020). As previously elucidated, the selected databases offer superior data quality and enable search query refinement within the search string. These databases are widely recognized as leading sources of analytical research data and scholarly citations (Chavarro et al., 2018; Li et al., 2018).

Online searches and retrievals were conducted on the Scopus and Web of Science core collections. The methodology employed for database search and retrieval aimed to comprehensively capture search items in a robust and comprehensive manner. Subsequently, a two-phase filtering process was implemented to retain the most pertinent outcomes. Documents such as book chapters, conference papers, textbooks, and editorials were excluded from consideration during the screening process. The initial search yielded a total of 601 items, encompassing a timeframe of 22 years (from 2000 to 2022). Distinct search strings (detailed in Tables 4.5 and 4.6) were employed to access pertinent materials from the Scopus and WoS databases, respectively.

Table 4.6: Search results from Scopus database (as of 18/08/2022).

No.	**Integrated search query**	**No. records**
1	TITLE-ABS-KEY((“Building Information Modelling” OR “Building Information Modelling” OR “Building Information Model” OR “BIM”) AND (“Bridge” OR “Dam” OR “Highway” OR “Major construction project” OR “Major project” OR “Railway” OR “Megaproject delivery” OR “Mega construction project” OR “Mega project” OR “Megaproject” OR “Airport” OR “Underground” OR “Tube” OR “High Speed Rail” OR “Olympics” OR “Olympic stadium” OR “National stadium” OR “Urban regeneration” OR “Urban redevelopment” OR “Transit” OR “Water infrastructure” OR “Waste water infrastructure”)) AND (LIMIT-TO (DOCTYPE”,ar”)) AND (EXCLUDE (PUBYEAR,1992)) AND (LIMIT-TO (LANGUAGE”,English”)) AND (LIMIT-TO (SUBJAREA”,ENGI”) OR LIMIT-TO (SUBJAREA”,BUSI”) OR LIMIT-TO (SUBJAREA”,ENVI”) OR LIMIT-TO (SUBJAREA”,COMP”) OR LIMIT-TO (SUBJAREA”,ENER”) OR LIMIT-TO (SUBJAREA”,MULT”))	412

Table 4.7: Search results from the Web of Science (WoS) database (as of 18/02/2022).

No.	Integrated search query	No. records
1	(“building information modelling” OR “building information modelling” OR “building information model” OR “BIM”) AND (“Bridge” OR “Dam” OR “Highway” OR “Major construction project” OR “Major project” OR “Railway” OR “Megaproject delivery” OR “Mega construction project” OR “Mega project” OR “Megaproject” OR “Airport” OR “Underground” OR “Tube” OR “High Speed Rail” OR “Olympics” OR “Olympic stadium” OR “National stadium” OR “Urban regeneration” OR “Urban redevelopment” OR “Transit” OR “Water infrastructure” OR “Waste water infrastructure”) (Topic) and Review Articles (Exclude – Document Types) and Articles (Document Types) and Engineering Civil or Construction Building Technology or Engineering Industrial or Engineering Multidisciplinary or Management or Architecture (Web of Science Categories) and Science Citation Index Expanded (SCI-EXPANDED) or Social Sciences Citation Index (SSCI) (Web of Science Index) and English (Languages)	189

4.4.1.2 Transforming the Data

After retrieving the '.bib' files from each database search, the researcher proceeded to convert them into a bibliometric format. For this purpose, the researcher utilized the 'biblioshiny()' command within the RStudio platform, which is made accessible through the 'biblioshiny()' function present in the Bibliometrix package (Aria and Cuccurullo, 2017). This approach leverages Biblioshiny to enable researchers who might not be familiar with RStudio to undertake this stage. The file conversion process remains consistent for each file and requires only repetition. The researcher brought in the '.bib' file into the Bibliometrix software by selecting the 'Import raw files' option from the 'Import or Load files' menu, which can be found in the 'Data' section. Following this, the researcher proceeded to save the file in 'Excel' format using the 'Export a bibliometrix file' choice within the same interface. To simplify the process, each file was designated as 'Scopus.xlsx' and 'WoS.xlsx'. As a result, two distinct Excel files were generated, both having the same tag fields in the initial row, but with information organized in different columns. Additionally, it was observed that specific supplementary details, such as URLs to publications, might be absent from the WoS dataset, potentially impacting the bibliometric analysis.

4.4.1.3 Merge the Data

Once all the datasets were standardized to the same format, they could be integrated. In this scenario, both Excel files were opened, and their windows were arranged horizontally for simultaneous viewing, enabling the researcher to have both files displayed on the screen in a stacked manner. Subsequently, the 'Scopus.xlsx' file was designated as the master file for merging with the Web of Science data. To ensure relevance to the study, any columns containing extraneous information in the master file were excluded. In the second file, 'WoS.xlsx', columns were reordered to align with the sequence of columns in the master file after the removal of unnecessary data was carried out manually. The procedure of copying and pasting content from one Excel file into the other was followed to generate a merged dataset. Finally, Excel's 'Remove Duplicates' feature was employed to eliminate any duplicated entries within the combined dataset.’

4.4.2 Text Analysis Process

4.4.2.1 ‘Trend Analysis

Upon concluding the data collection phase, the researcher merged the outcomes of both Scopus and WoS searches into a single Excel file. This consolidated file was subsequently imported into Bibliometrix to facilitate the execution of bibliometric analyses. The refined outcomes underwent co-occurrence and frequency analyses, leveraging specific attributes of the published materials such as author associations, publication dates, publication types, journal sources, and country affiliations, among other factors. Subsequently, keyword co-occurrence analyses were conducted to delve into potential research directions, knowledge gaps, and trends. The objective of these analyses was to evaluate BIM research within the context of major construction project delivery.

4.4.2.2 Bibliometric and TM Analysis

Semi-structured, structured, and heterogeneous data types possess significant relevance for analyses and procedures within the Architecture, Engineering, and Construction (AEC) sector, particularly in the context of BIM. The exploration of knowledge can be approached through a variety of methods, encompassing mathematical, nonmathematical, and inductive approaches. The findings of this study carry implications across a range of fields, including data maintenance, decision support, information management, and query optimization. As individuals interact with collections of documents, they utilize a variety of analytical tools to unveil and investigate patterns of interest (Feldman and Sanger, 2006). Differing from conventional data mining, TM focuses on analyzing and modelling unstructured natural language text from sources such as news articles, scientific research papers, and medical documents. According to Grobelnik et al. (2002), this technology integrates NLP, pattern categorization, ML, and statistical techniques, forming a comprehensive solution.

It is becoming increasingly difficult, to extract high-quality information from the vast amounts of data that are being generated in the network as a result of the rise of Big Data (Cai and Zhu, 2015; Jones et al., 2004). As the most frequent kind of web data, unstructured text has a wealth of information and trends that can be used to improve management practices (Li and Wu, 2010). This has led to numerous research that explore the relationship between management and TM. Scholars and management practitioners can better understand how text-mining techniques can be integrated into management operations by obtaining the theoretical underpinning of this literature. To determine which variables have the biggest influence on the forecasted outcome, it is necessary to conduct an analysis. It would take an inordinate amount of time to browse and analyse a dataset with hundreds of fields. At this point, a potent approach to assist researchers in completing these tasks must be selected.

As a typical strategy to synthesize research results, bibliometrics employs mathematical and statistical approaches to examine books or other communication media in a certain field in order to uncover key players and publications in the field of research, as well as to identify research themes and trends (Pritchard, 1969). While there have been several bibliometric studies in the academic community looking at the use of TM techniques in various fields (Wen et al., 2021), none have looked at the application of TM in BIM pervasive major project delivery to date. To address this, this work collects relevant literature from the Scopus and WoS database and carries out bibliometric analysis by identifying the most influential journals, countries, authors, and literature in the study of using TM in BIM pervasive major project delivery, as well as the revealing related evolutionary trends.

According to Chellappandi and Vijayakumar (2018), metric studies, which are based on standardized measurement techniques, capture the current study direction of information and library studies. The measurement of features of documents, formats, books, texts, and other information is referred to as bibliometrics. As per Pritchard (1969), bibliometrics uses sorting, statistical and mathematical approaches to various forms of communication media. While bibliographic methods are traditional in information sciences and library studies, a diverse range of fields employ them to scrutinize research trends and implications within their respective domains. Subsequently, the field of metric studies has expanded to cover various disciplines, incorporating areas like bibliometrics, scientometrics, altmetrics, cybermetrics, informetrics, and webometrics, among others. Categorization strategies offer numerous examples of this, such as capture, citation methods (e.g., Scopus and Web of Science citations), downloads, mentions in social media (e.g., Twitter and blogs), and methods like keywords or bookmarking. An important factor in the recent explosion of metric studies is the availability of sophisticated statistical techniques, databases, and computer technologies that make it possible to conduct these analyses (Sajovic et al., 2018).

Sajovic et al. (2018) noted that the examination of visual attributes concerning differences and similarities within data collected from single or multiple sources, along with relationships like co-authors, bibliographic coupling, citations, data, and publication source timelines, can be represented in bibliometric investigations through the application of techniques and algorithms for science mapping. This study used these approaches to assess research trends in BIM pervasive major project delivery to help create research timelines and benchmarks for future studies. In addition, bibliographic maps were employed to model the domain’s physical structure and characteristics. Knowledge domains in construction management research such as intellectual structure and BIM evolution, green buildings; construction safety management; construction waste and demolitions; and the use of contingency theory in project management are all examples of bibliometric analyses that are common in construction management research (Hanisch and Wald, 2012; Liang et al., 2020; Olawumi et al., 2017; Oshodi et al., 2020; Wu et al., 2019; Zhao et al., 2019).

Approaches used for this study are listed below in detail:

- Examinations of author, region/country, source, publication type, and organizational publication frequencies spanning the years 2000 to 2022, depicted in Figure 4.1.
- Investigations into co-authorship within the study's corpus, encompassing insights into authors' co-occurrences and collaborative networks.
- The associations and interdependencies among relevant terms and trends within the context of BIM pervasive major project delivery literature can be discovered through the co-occurrence of keywords analysis. The co-occurrence of terms was limited to the years 2014–2022, to uncover recent trends and other related exclusion criteria.
- Using the entire text of research papers from the Scopus and WoS databases, the content of the research body reported in the articles can be mapped and mined. In addition to keywords, relevant noun phrases and nouns are also identified, providing better insights into the growing research.

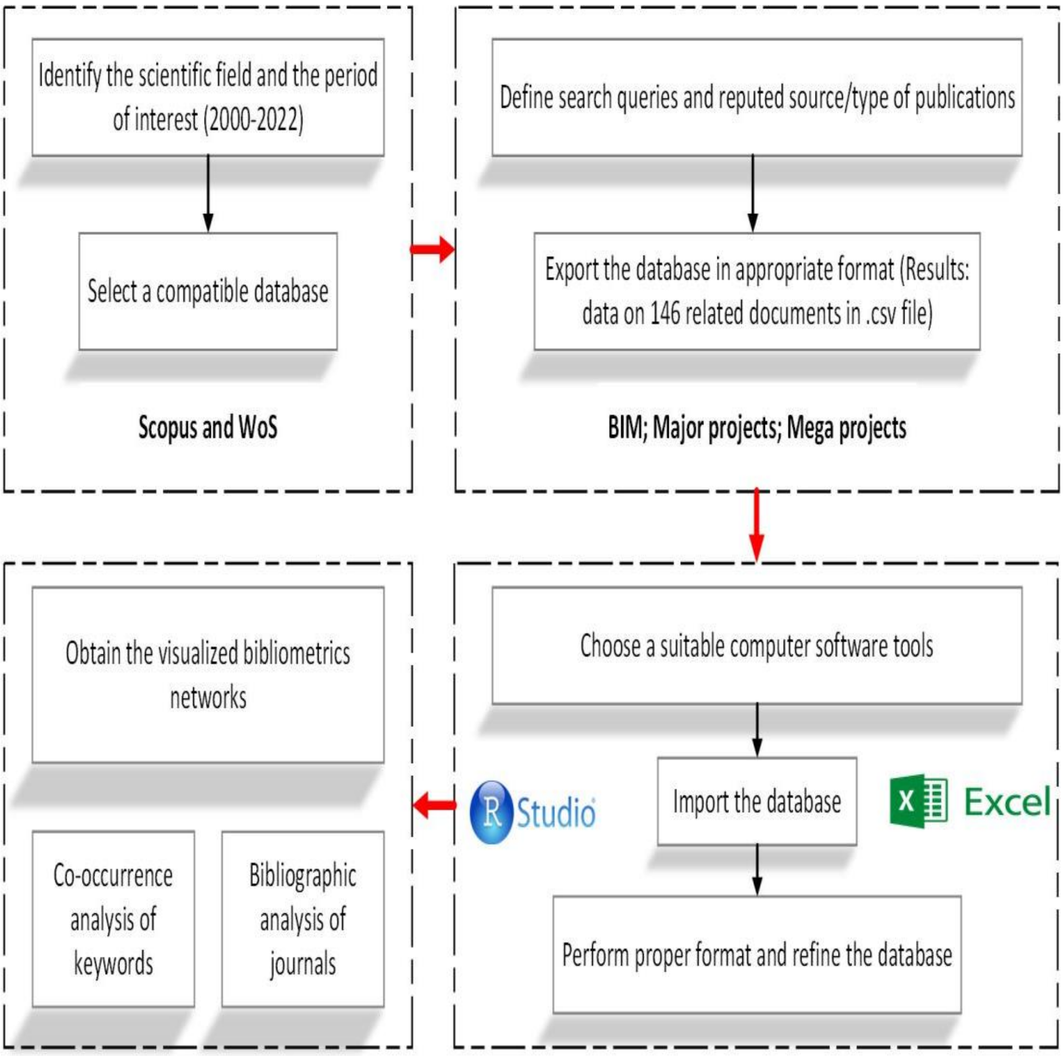


Figure 4.1: Process of bibliometric analysis

Source: Wrigley et al. (2019)

4.4.2.3 Selecting the Analytical Tool

Increased tools and algorithms originating from network theory and reinforced by the improved usage of computers and internet are used to highlight the strength of social network impacts on research trends and in research (Smiraglia, 2015). Network analysis and visualization have been facilitated by an array of tools including Citespace, Pajek, Biblioshiny, Gephi, nodeXL, VOSviewer, and various others. In this study, the 'biblioshiny()' command from RStudio, accessible via the Bibliometrix package's 'biblioshiny()' function, was employed. The industry's researchers are progressively acknowledging the utility of RStudio software, harnessing its capabilities especially for graphical (Aria and Cuccurullo, 2017) and metadata metric investigations, among other use cases.

The data utilized in this experiment were collected from the Scopus and WoS databases. The research primarily cantered on analyzing the co-occurrence of keywords and conducting co-authorship analyses. To achieve this, the RStudio's 'biblioshiny()' command was employed. Labels and circles are used to represent the elements in the visualization network in the biblioshiny results. The number of links between them represents the strength of the association between objects. A cluster analysis of keyword co-citation is performed using the lin-log normalization technique, which determines the colouring of the items and the cluster to which they belong while doing a cluster analysis of keyword co-citation.'

4.4.3 Frequency Analysis

4.4.3.1 'Journal Publications

Based on the data in the Scopus and WoS database, 146 different publications were found. Table 4.7 presents the top 10 most frequently cited publications along with the number of published documents and the recorded number of citations. According to Guo et al. (2019), when an article or document is cited multiple times, it is said to have a high impact on the scientific community. In the Journal of *Automation in Construction*, sixteen papers gathered a total of 548 citations, whereas one paper in the *Journal of Advanced Engineering Informatics* acquired 73 citations. The *Journal of Engineering Construction and Architectural Management* exhibited the lowest figures, with five articles and four citations.

Table 4.8: The 10 most often cited publications

No.	Journals	Documents	Citations
1	Advances in Civil Engineering	8	5
2	Applied Sciences	5	25
3	Automation in Construction	16	548
4	Journal of Construction Engineering and Management	4	31
5	Journal Of Computing in Civil Engineering	6	67
6	KSCE Journal of Civil Engineering	5	26
7	Engineering Construction and Architectural Management	5	4
8	Advanced Engineering Informatics	1	73
9	Tunnelling and Underground Space Technology	9	70
10	Sustainability	9	34

4.4.3.2 Distribution across Regions and Countries

Figure 4.2 illustrates the representation of publication activity for BIM research in the context of major construction project delivery across different countries, with a total of 146 published documents from five continents:

- Asia (n = 67) (China 46, Korea 15, Turkey 3, Pakistan 2, Hong Kong 1, India 1, Thailand 1).
- North America (n = 16) (USA 13, Canada 3).
- Europe (37) (UK 19, Italy 6, Germany 4, Norway 3, Netherlands 1, Portugal 1, Serbia 1, Spain 1, Sweden 1).
- Australasia (n = 7) (Australia 6, New Zealand 1).
- Africa (n = 4) (Egypt 3, Morocco 1)

No study was found from a South American country; this may reflect publication bias in this search, as it was limited to studies in English (it is possible that related South American studies were conducted and published in Spanish and Portuguese). Although this list is not comprehensive, it shows the continents and countries where most BIM pervasive major project delivery research is currently being published. The limited findings from Africa reaffirm Saka and Chan's (2019; 2020) findings on BIM use and adoption in Africa. Although most African countries use BIM technology, major usage and adoption challenges are still prevalent (Hamma-adama and Kouider, 2019; Mtya and Windapo, 2019; Nasila and Cloete, 2018; Olanrewaju et al., 2020). The BIM paradigms aimed at adaptation in Africa are categorized into three stages: awareness, implementation, and improvement. As depicted in Figure 4.3, the majority of published BIM research pertaining to major construction project delivery across the globe is concentrated in countries situated in Asia, Australia, Europe, New Zealand, and North America.

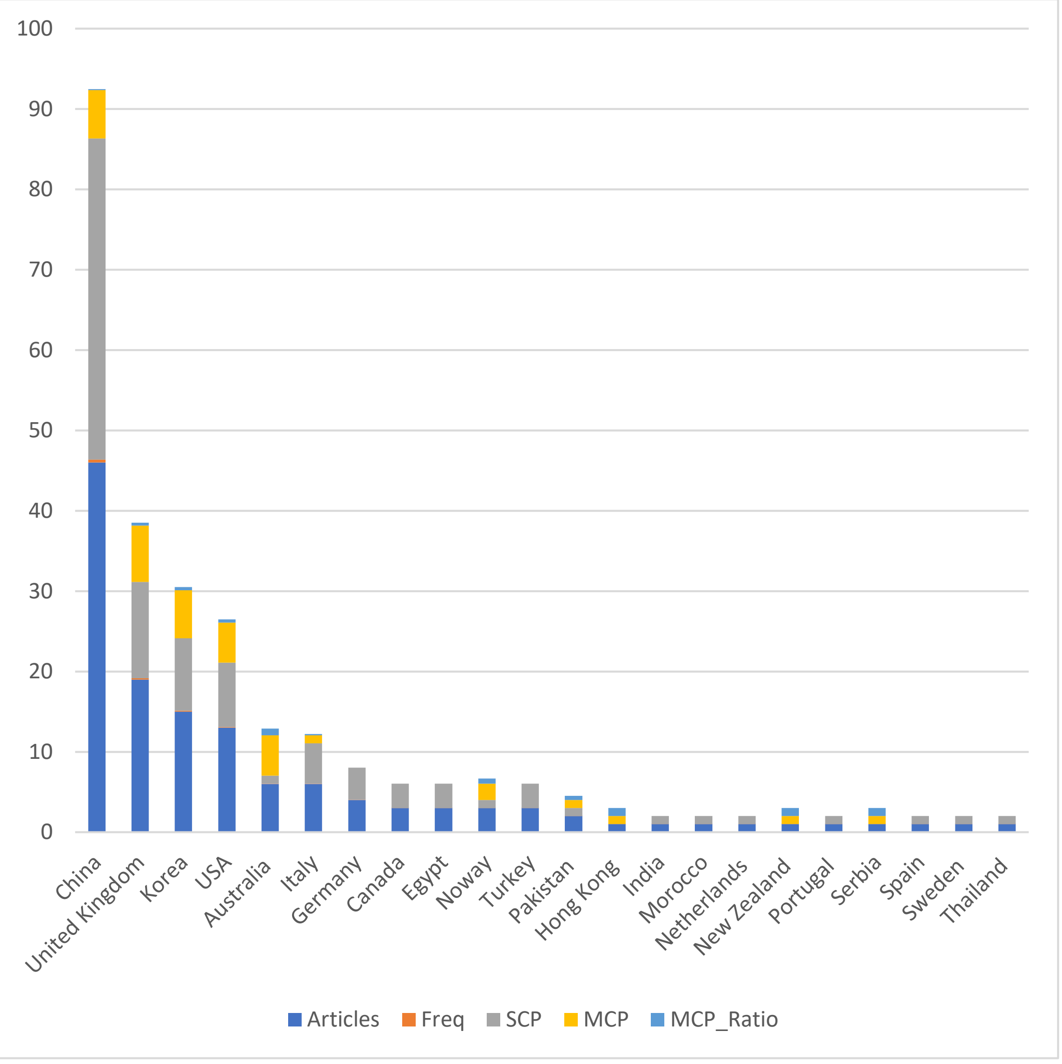


Figure 4.2: The distribution of countries' publication activity

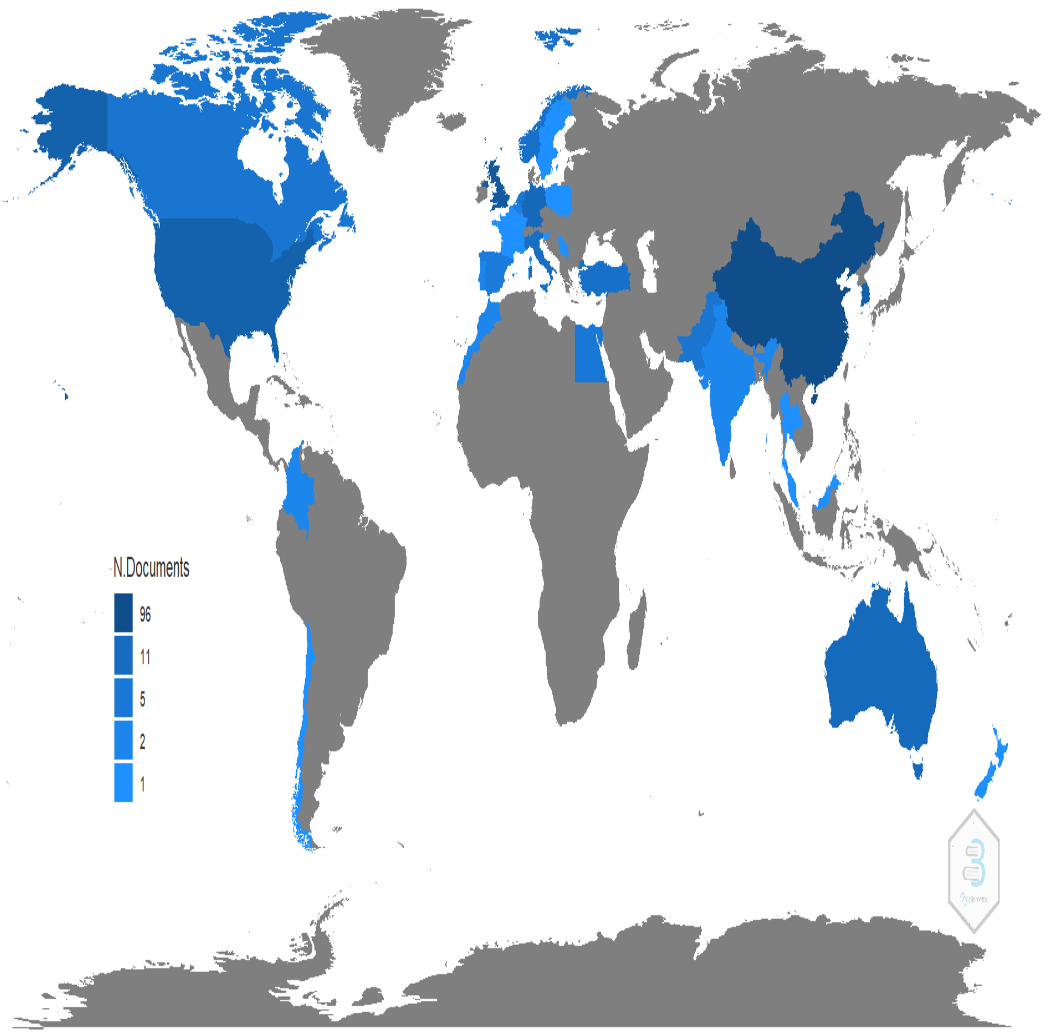


Figure 4.3: Global distribution of visible published BIM studies

4.4.3.3 Co-Authorship Network and Publication per Author

The analysis of notable organizations, scientific collaboration patterns, individual researchers, or countries can be substantially enhanced through the utilization of a co-authorship network (Fonseca et al., 2016). As illustrated in Figure 4.4, the network showcases authors who have contributed to research and trends related to BIM pervasive major project delivery. The software functionality was deemed notable if it was associated with a minimum of five documents, each with at least 10 citations. Overall, the dataset comprised 442 authors across 146 documents, with 45 authors fulfilling the criteria for significance. Table 4.9 presents the 20 most prolific authors within this study's domain. The evaluation of research impact involved utilizing Total Citation (TC) scores, TC per year, and normalized TC metrics.

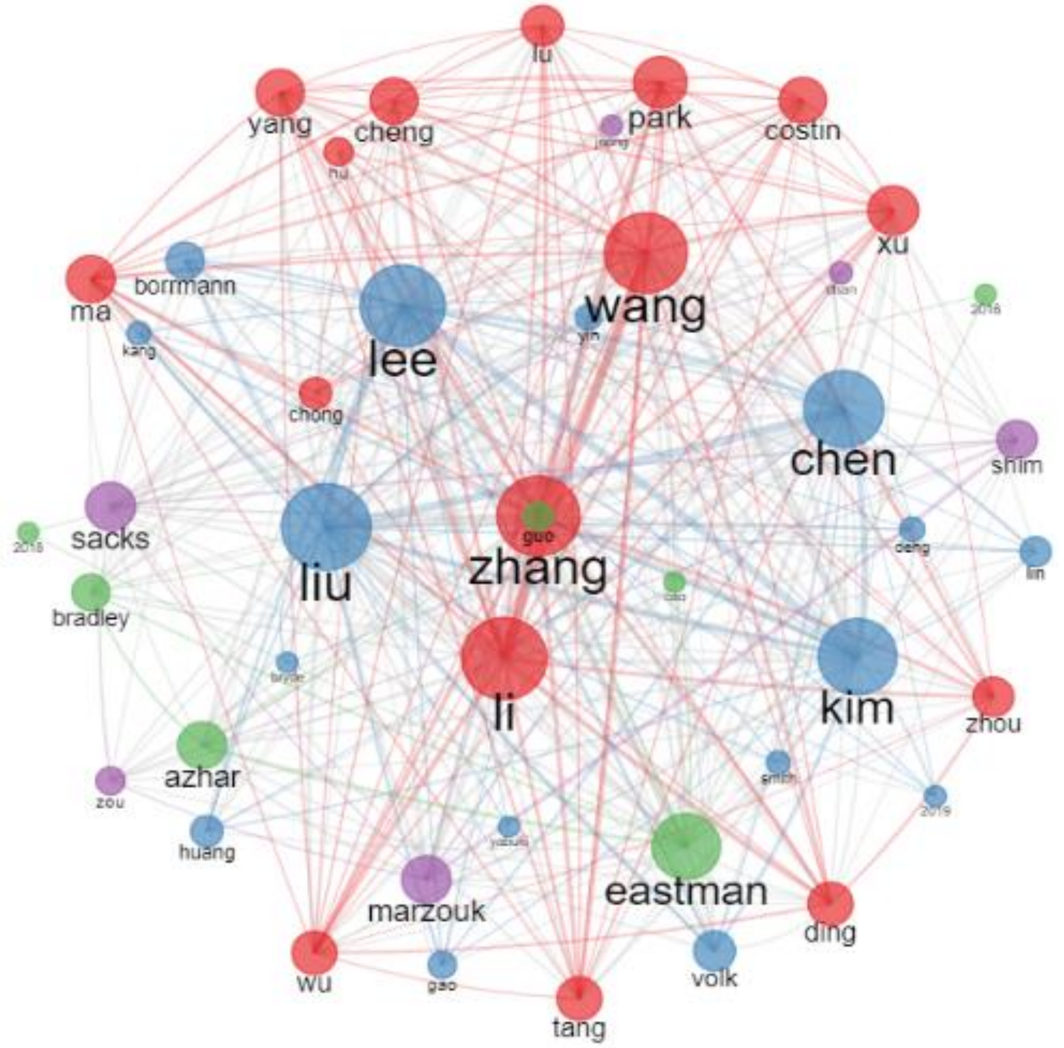


Figure 4.4: Network of authors who contributed prolifically to BIM research and trends

Table 4.9: The most 20 prolific authors in this study domain

No.	Authors and Year	Total Citations	TC per year	Normalized TC
1	M. Marzouk, 2014	74	8.22	2.8462
2	B. Mcguire, 2016	61	8.71	2.6237
3	L. Zhang, 2016	60	8.571	2.5806
4	B. Fanning, 2015	49	6.125	2.2529
5	S. Kaewunruen, 2019	45	11.25	3.4615
6	F. Tang, 2020	40	13.333	4.9485
7	PC. Lee, 2018	36	7.2	1.9412
8	Y. Zou, 2016	35	5	1.5054
9	Q. Wang, 2018	34	6.8	1.8333
10	Ped. Love, 2017	32	5.333	1.8462
11	F. Tang, 2020	32	10.667	3.9588
12	Z. Hu, 2016	31	4.429	1.3333
13	M. Li, 2018	31	6.2	1.6716
14	M. Wang, 2019	30	7.5	2.3077
15	CS. Shim, 2017	28	4.667	1.6154
16	X. Yin, 2020	28	9.333	3.4639
17	Z. Aziz, 2017	27	4.5	1.5577
18	J. Delgado, 2018	26	5.2	1.402
19	L. Yang, 2020	25	8.333	3.0928
20	M. Marzouk, 2014	24	2.667	0.9231

A comprehensive count of 442 authors associated with 146 documents successfully met the established criteria, encompassing both contributing and lead authors. The 'biblioshiny' command, harnessed within RStudio software's analytical algorithms, necessitated a minimum of five documents and ten citations for each author.

Consequently, 67 authors fulfilled this stipulation. Noteworthy among these are the 20 researchers who have made significant scientific contributions and wielded substantial influence within the realm of BIM pervasive major project delivery research, as highlighted in Table 4.9. Notably, the following authors emerged as the most prolific contributors in the subject of study:

- “B. Fanning (49 TC, 6.12 TC per year, 2.25 normalized TC)”
- “S. Kaewunruen (45 TC, 11.25 TC per year, 3.46 normalized TC)”
- “M. Marzouk (74 TC, 8.22 TC per year, 2.84 normalized TC)”
- “B. Mcguire (61 TC, 8.71 TC per year, 2.62 normalized TC)”
- “L. Zhang (60 TC, 8.57 TC per year, 2.58 normalized TC)”

Another notable researcher in the field is F. Tang (40 TC, 13.33 TC per year, 4.94 normalized TC). Leveraging the comprehensive array of connections and their associated strengths, as previously established, the study delved into the overlay and network visualizations showcased in Figure 4.6. In this context, a connection signifies collaborative engagement among authors, while the aggregate link strength within the co-authorship networks (Van Eck and Waltman, 2010) denotes the interconnectedness between sources.

Figure 4.4 illustrates the network visualization, revealing the identification of 20 co-authorship clusters. Within these clusters, a total of 45 network objects were identified, contributing to an aggregate connection strength of 218 among the authors. Notably, 29 connections were observed among authors within the research domain. Noteworthy collaborative clusters of authors, such as F. Tang, M. Marzouk, B. Mcguire, B. Fanning, and S. Kaewunruen, have displayed active research collaboration. The research output has displayed a gradual increase from 2014 to 2022, evident from the publication timelines outlined in Table 4.10.

Table 4.10: Publication timetables from 2014-2022

Descriptions	Results
Main information about data	
Timespan	2014-2022
Sources (journals, books, etc.)	66
Documents	146
Average years from publication	2.79
Average citations per documents	10.23
Average citations per year per doc	2.289
References	6123
Document types	
Article	146
Document contents	
Keywords plus (ID)	996
Author’s keywords (DE)	475
Authors	
Authors	442
Author appearances	541
Authors of single-authored documents	6
Authors of multi-authored documents	436
Authors collaboration	
Single-authored documents	6
Documents per author	0.33
Authors per document	3.03
Co-authors per documents	3.71
Collaboration index	3.11

4.4.3.4 An Organization's Or Institution's Output of Publications

This section examines the extent of research projects and publications conducted at the university and organizational levels. The inclusion of materials was guided by specific research criteria. A total of 197 organizations and universities were identified, and Table 4.11 features 25 documents published by diverse institutions and universities since 2014.

Table 4.11: Organizational output of publications

No.	Affiliations	Documents
1	Bauhaus-University Weimar	3
2	Beijing University of Technology	3
3	Central South University	3
4	China University of Mining and Technology	3
5	Chongqing Jiaotong University	3
6	Curtin University	3
7	Hanyang University	3
8	Harbin Institute of Technology	3
9	Huazhong University of Science and Technology	3
10	Newcastle University	3
11	Syracuse University	3
12	The Hong Kong University of Science and Technology	3
13	The University of Nottingham	3
14	Tongji University	3
15	Chung-ang University	4
16	Colorado State University	4
17	Shanghai University	4
18	Southeast University	4
19	Southeast University	6
20	Tianjin University	4
21	University of Birmingham	6
22	University of Liverpool	4
23	Xian University Architecture and Technology	4
24	Yonsei University	4
25	Bauhaus-University Weimar	3

4.4.3.5 Final Publication Output Step

In terms of prevalence, the most prominent journal sources were primarily oriented towards construction and building technology, environmental engineering, civil engineering, multi-disciplinary engineering, industrial engineering, and computer science. To facilitate a more comprehensive examination of themes related to BIM pervasive project delivery, the search was refined to align with the specific focus. Following the removal of book chapters and summaries from the gathered articles in both the Scopus and WoS collections, a fresh set of filtering criteria was employed to organize the remaining 601 papers. Applying these criteria revealed that no papers were published prior to 2014. Consequently, to discern research trends, maintaining the study's designated time frame was imperative. The subsequent section briefly outlines the additional filtering criteria employed in the ultimate selection of the extracted documents.

- Articles published before 2014 (i.e., from 2000 to 2013) had to be removed from the study, because the focus was on the most recent research in BIM pervasive major project delivery. Consequently, articles from 2014 to 2022 were evaluated.
- With BIM being a computing phenomenon, literature and developer options on computing and mathematical derivations were not included in the considered research results. Journals of computing and related literature were screened for exclusion.
- Articles in languages other than English were omitted, and papers lacking a minimum of five distinct keywords were manually excluded by the researcher due to their limited contribution to the expanding knowledge in the BIM field.

After applying these filtering criteria, a total of 146 bibliographic documents were identified as meeting the selection criteria for further investigation. All the materials (100 percent) used in this study consisted of articles published between 2014 and 2022 (Table 4.10). The second phase of filtering involved the utilization of relevant publication and co-occurrence criteria.

4.4.3.6 Keyword Network Formed by Co-occurrences

The core essence of a publication's content is mirrored through its keywords, which encompass concepts, nouns, and essential phrases (Xiang et al., 2017). The term "co-occurrence" pertains to the nearness or presence of similar keywords. As outlined by Lozano et al. (2019), the proximity of keywords in co-occurrence determines the likelihood of their appearance together. Employing the 'biblioshiny' command within RStudio software, a co-occurrence network was constructed using a total of 730 significant keywords, as visualized in Figure 4.7. A minimum inclusion threshold of 5 occurrences was used to select keywords, and this criterion was met by forty-five (45) out of the 730 key terms, forming four (4) clusters. Figure 4.7 illustrates the identified clusters. The ensuing discussion will delve into the clusters of keywords, revealing that certain frequently associated terms with BIM may now hold less weight, as the focus shifts towards emerging trends and innovations.

4.4.3.7 Cluster 1

Cluster 1, depicted in red in Figure 4.5, encompasses a total of 22 associated terms. The keywords within this cluster comprise architectural design, automation, three-dimensional computer graphics, risk assessment, safety engineering, railroads, efficiency, integration, light rail transit, structural design, and underground construction. This cluster revolves around pre-design considerations for BIM, offering an optimization plan for the project team's proposed project. Therefore, this cluster can also be referred to as 'BIM Performance Indicators' (BIM PI).

4.4.3.8 Cluster 2

Highlighted in blue, another cluster is tied to 11 keywords, featuring terms like information theory, maintenance, information systems, roads, and streets. This cluster centers around 'Information System' and is encapsulated as 'BIM IS' for brevity.

4.4.3.9 Cluster 3

Marked in green, this cluster comprises 5 keywords. For example, the cluster includes unusual terms like industry foundation classes – IFC, office buildings, and tunnels. BIM Industry Foundation Classes are summarised in this cluster, ‘BIM IFC’.

4.4.3.10 Cluster 4

Presented in a gentle shade of yellow, this cluster encompasses 7 distinct terms. Keywords within this cluster encompass information management, decision-making, lifecycle, visualization, data visualization, and inspection. This cluster is cantered around 'BIM Solution Content and Applications' ‘BIM SCA’.

The results of this keyword co-occurrence analysis highlight the limited research focus on BIM pervasive major project delivery. Emerging research and communities are observed in various areas, including information theory, decision-making, lifecycle, efficiency, automation, three-dimensional computer graphics, risk assessment, safety engineering, and Industry Foundation Classes (IFC), an open-source neutral specification guide not controlled by any vendor (Ghaffarianhoseini et al., 2017). It is noted that there is room for standardization, enhancement, and application of BIM to cater to the needs of a growing client base. According to Ahuja et al. (2017) and Olawumi et al. (2017), the construction industry stands to gain from advancements in BIM technology.’

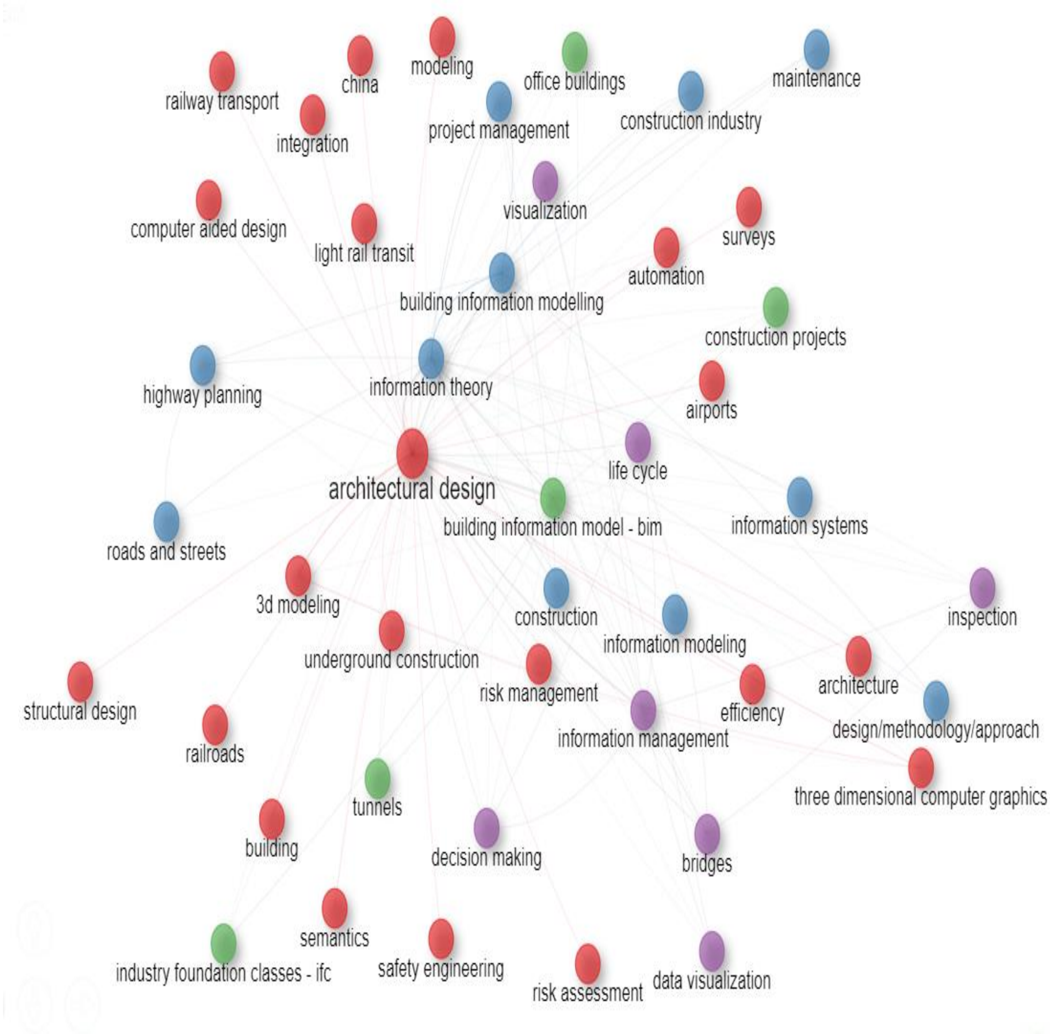


Figure 4.5: Co-occurrence network keywords

4.4.4 Trends in BIM Pervasive Major Project Delivery

4.4.4.1 Overview

Between 2014 and 2022, the dimension, focus, and volume of BIM studies underwent significant improvements. Notably, the visualization of subjects that characterized the initial year of the study, 2014, was particularly intriguing. As the focus shifted from exploratory topics like infrastructure planning and atmospheric temperature, the number of publications increased, and more specific themes emerged. For instance, in 2015, attention turned to subjects such as the 3D underground cadastral system, accuracy assessment, cost reduction, geometrical modelling, and information technology. The subsequent years witnessed a rising interest in transportation-related matters, including roads, streets, bridge inspection, laser scanning, and transportation infrastructure. The years 2017 and 2018 saw growing popularity in subjects like modelling, management, railway transport, information theory, information management, and decision-making within the realm of BIM research.

As illustrated in Figure 4.6, there is a noticeable rise in the utilization of BIM in various domains. Notably, 3D computer graphics, data visualization, underground construction, and airports witnessed increased prominence in the early months of 2020. Subsequently, in the years 2021 and 2022, topics like metro stations, industry foundation class, and augmented reality gained traction within the research landscape. These trends showcased in the analysis highlight the progression of research and its interconnectedness with the sphere of BIM pervasive major project delivery.

The outcomes of this study underline a significant transformation in the exploration of BIM. Nevertheless, it is essential to underscore that certain aspects of BIM, such as its utilization for performance enhancement, productivity enhancement, and governance implementation, remain relatively unaddressed.

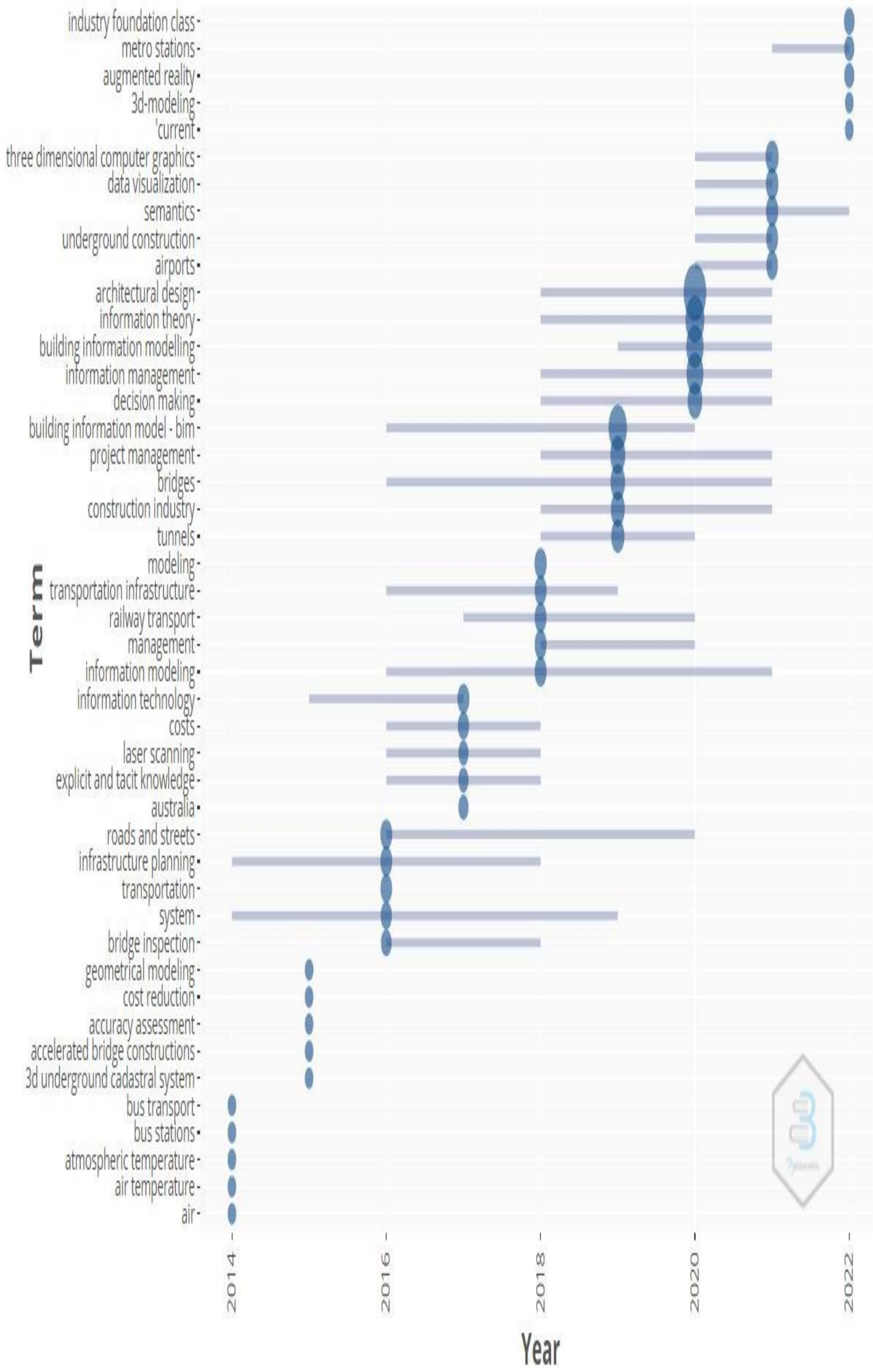


Figure 4.6: Trend analysis results

4.4.4.2 Text Mapping

Text mapping was carried out using RStudio, where the full-text content of the 146 documents was scanned, and text extraction from the corpus was performed based on the occurrences and significance of active nouns. The software assessed term relevance, requiring a minimum of 20 occurrences for inclusion. Among the 52,065 words in the list, 808 words met this criterion. Filtering options were applied, including the removal of word variations and their acronyms, to ascertain relevance. Table 4.12 provides the list of the top 20 most frequently occurring words in the analyzed full-text articles. In addition, Figure 4.9 depicts a word cloud depicting the most active words relevant to TM activity. The software was programmed to generate 200 words, with the most active words focusing on the construction sector and project requirements. Words in orange identified 18 items such as technology, software, data, monitoring, IFC, maintenance, etc. Red shading is used to emphasize the design aspect, while blue highlights represent the system, process, and management. The most prominently featured terms are marked in green, denoting their relevance to the application of BIM in significant construction project delivery. These terms encompass concepts such as performance, risks, environment, simulation, development, operation, implementation, and scheduling.

Table 4.12: Top 20 words in analysed full text articles

No.	Words	Frequencies		No.	Words	Frequencies
1	Data	6224		11	Structure	1374
2	Design	3618		12	Proposed	1328
3	System	2956		13	IFC	1325
4	Management	2929		14	Method	1313
5	Process	2178		15	Cost	1276
6	Models	1815		16	Safety	1240
7	Risk	1689		17	Software	1212
8	Analysis	1608		18	Technology	1203
9	Time	1472		19	Level	1160
10	Maintenance	1429		20	Structural	1151

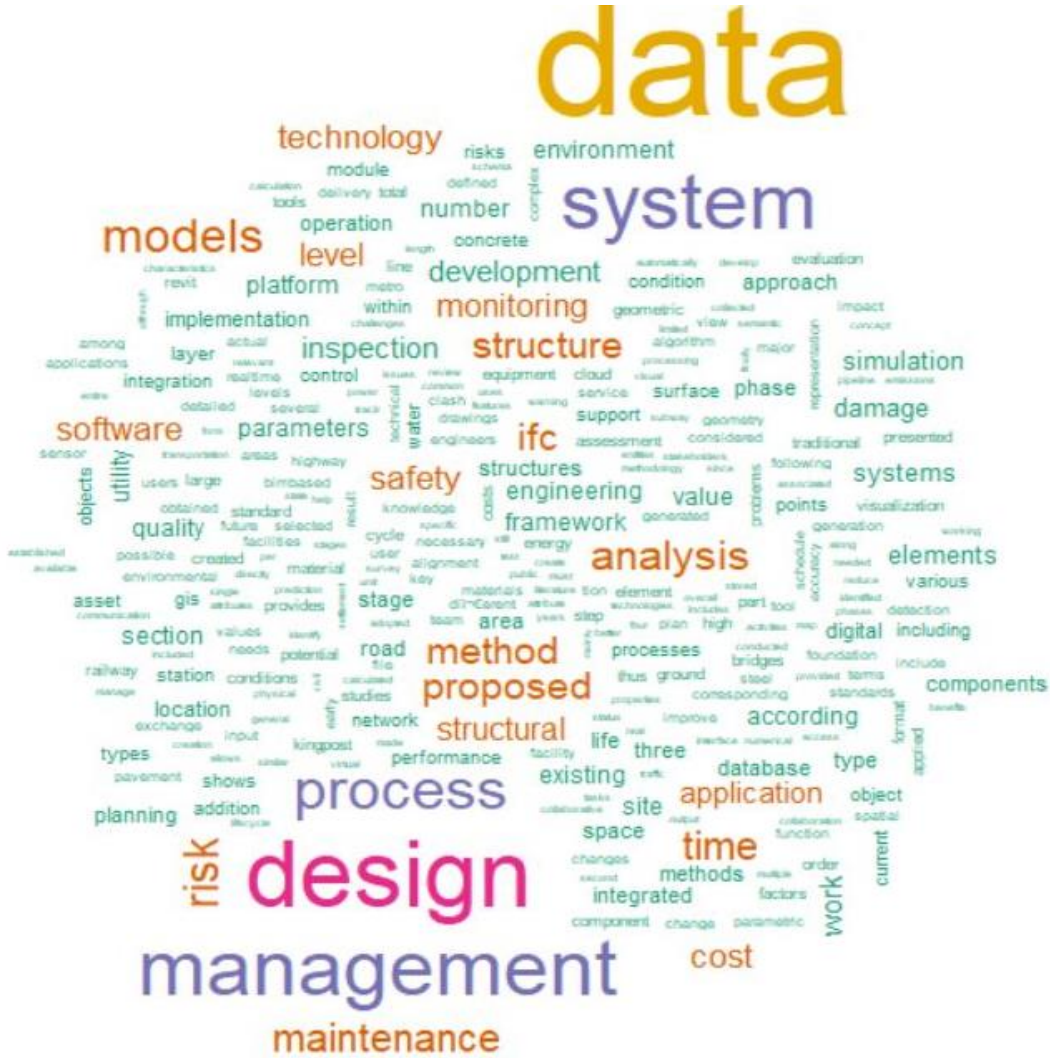


Figure 4.9: Word cloud

4.4.4.3 Knowledge Graph

The development of the BIMTAPE Framework involves integrating TM techniques into the landscape of BIM. The "Knowledge Graph" technique, exemplified through the graph derived from BIM text analysis, offers a dynamic visualisation showcasing how various BIM terms influence one another within the BIMTAPE Framework.

Each node in the knowledge graph represents a key technical term with regard to the use of BIM, such as Data, Design, System, etc., and the edges signify connections and associations between these terms based on their co-occurrence or thematic relevance. This nuanced representation, resulting from the application of TM, becomes a crucial aspect of the BIMTAPE Framework, emphasising the interdependencies that shape BIM processes.

The knowledge graph presented in Figure 4.10, derived from text analysis on BIM documentation, provides a detailed representation of key terms and their relationships within the BIM domain. Each node in the graph corresponds to a specific term. Developed using Python, this graph serves as a visual representation of the influence each BIM oriented term has on others, and offers a dynamic and interconnected view of technical concepts and processes for BIM pervasive construction project delivery. Importantly, the graph intentionally includes only twenty terms (see Table 4.12), carefully chosen based on their frequency in the text. These terms represent the key technical term with regard to the use of BIM. The deliberate selection aims to make it easier to compare these key technical terms used in BIM with those in the Periodic Table of BIM (Mordue, 2016). This approach enhances the understanding of how these two sets of terms relate to each other.

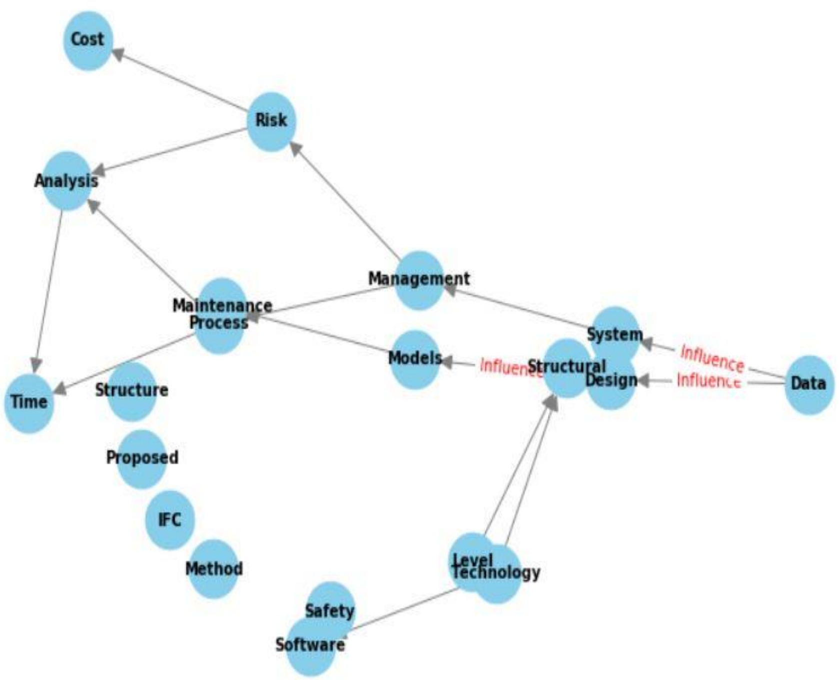


Figure 4.10: knowledge graph of key technical terms of BIM

When connecting the knowledge graph with the Periodic Table of BIM (Mordue, 2016), the influence of each word on others takes centre stage. The Periodic Table's structured categorisation aligns with the dynamic insights provided by the knowledge graph, creating a cohesive narrative of how various BIM elements influence and interact with one another. This integration serves to enhance the BIMTAPE Framework, as the BIMKG emerges as a new theoretical method for supplementation and complementation beyond the Periodic Table of BIM (Mordue, 2016). By visualising the dynamics of BIM implementation, this innovative approach provides substantial support for decision-making and knowledge engineering in BIM pervasive project management, particularly within major project delivery. The visual and structured representation of each term's influence on the overall BIM landscape contributes to a deeper understanding of the interconnected nature of BIM elements and further strengthens the strategic insights offered by the BIMTAPE Framework.

For instance, terms such as Data, Design, and System (Table 4.13) identified through the TM, exhibit influence within the Periodic Table's classifications. Data aligns with the Strategy and Foundations domain, Design corresponds to both Foundations and Collaboration, and System resides within the Process category. This alignment offers a detailed understanding of how these key terms interconnect within the broader BIM technical domains. The structured representation in the table enhances comprehension of the intricate relationships between BIM elements, reinforcing the BIMTAPE Framework's objective of performance enhancement in major project delivery.

In the context of the provided BIM elements list, the specific relationship between the word Data and the BIM elements is not explicitly mentioned. However, based on common BIM terminology (McPartland, 2015-2018), Data is often considered foundational or fundamental to various aspects of the BIM process. In this case, it may be associated with the Foundation element, signifying its essential role as a basis or core component within the broader BIM strategy data.

The word Design in the context of BIM is typically associated with the Collaboration and Process domains. Design is a collaborative process involving various stakeholders, and it significantly contributes to the overall BIM process (El-Diraby et al., 2017). Therefore, it aligns with the Collaboration domain, emphasizing teamwork and joint efforts in the design phase. Additionally, Design is inherently linked to the Process domain, indicating its role in the systematic workflow and procedures within the BIM framework.

The word System in the context of BIM is commonly connected to the Process domain. In the BIM framework, the term System reflects the organized and systematic approach to managing and coordinating various elements and processes (Papadonikolaki, 2018). It implies a structured workflow and procedures within the BIM environment, aligning with the broader activities and methodologies outlined in the Process domain. The Process domain encompasses the systematic execution of tasks and activities throughout the lifecycle of a construction project, where the term System finds its relevance.

Table 4.14 systematically aligns prioritized key terms derived from the TA process with their strategic implications within the Strategy domain of the Periodic Table of BIM, as outlined by Mordue (2016). This strategic alignment reveals the nuanced connections between individual key terms and specific elements captured within the Strategy domain.

For example, the term Data is strategically associated with considerations of BIM strategy relating to data management, while Design corresponds to strategic factors of BIM design processes. This meticulous mapping enhances the understanding of the intricate relationships between key terms and strategic elements within the broader BIM landscape, providing valuable insights into their strategic implications and applications. It is imperative to acknowledge the foundational contribution of Mordue (2016) in formulating the Periodic Table of BIM, which serves as a theoretical framework guiding the structured organization of BIM elements across diverse technical domains.

Table 4.13: A reflection of prioritised technical essentials to the Periodic Table of BIM.

Priority	**Key terms from TA process**	**Technical Domains from the Periodic Table of BIM**								
		Strategy	Foundation	Collaboration	Process	People	Technology	Standards	Enabling tools	Resources
1	Data	✓	✓							
2	Design			✓	✓					
3	System				✓					
4	Management				✓	✓				
5	Process	✓			✓					
6	Models						✓			
7	Risk				✓					
8	Analysis				✓				✓	
9	Time									
10	Maintenance									✓
11	Structure		✓							
12	Proposed							✓		
13	IFC				✓					
14	Method				✓					
15	Cost				✓					
16	Safety						✓			
17	Software						✓			
18	Technology						✓	✓		
19	Level					✓				
20	Structural		✓							

Table 4.14: A reflection of prioritised key terms to technical domain on Strategy in the Periodic Table of BIM

Priority	**Key Terms from TA Process**	**Reflections to BIM Domain and Elements on Strategy** (Mordue, 2016)
1	Data	BIM strategy on data
2	Design	BIM strategy for design
3	System	BIM strategy on system development and/or adoption
4	Management	BIM strategy on management
5	Process	BIM strategy on process
6	Models	BIM strategy on model development and/or adoption
7	Risk	BIM strategy on risk management
8	Analysis	BIM strategy on analysis and evaluation
9	Time	BIM strategy on time management
10	Maintenance	BIM strategy on maintenance and lifecycle
11	Structure	BIM strategy on structural components and systems
12	Proposed	BIM strategy on proposed solutions and alternatives
13	IFC	BIM strategy on Industry Foundation Classes (IFC) integration
14	Method	BIM strategy on methods and procedures
15	Cost	BIM strategy on cost estimation and management
16	Safety	BIM strategy on safety planning and implementation
17	Software	BIM strategy on software applications and integration
18	Technology	BIM strategy on technology
19	Level	BIM strategy on information and data levels
20	Structural	BIM strategy on structural design and implementation

4.4.5 Lessons Learned from Experiment 1

This investigation revealed that emerging BIM publications and research directions are originating from regions such as Asia, the USA, Europe, New Zealand, and Australia, as previously mentioned. Notably, the study's analysis of BIM trends in Africa did not yield significant results, aligning with the conclusions of Saka and Chan (2020) and Olawumi et al. (2017), who highlighted the challenges African countries face in BIM adoption. However, the study underscores the changing landscape of the construction sector, where BIM is gradually being integrated into policy and practical implementation worldwide. This trend is propelled by the growing demand for environmentally sustainable and efficient construction practices.

The bibliometric analysis of scholarly journals in this study unveils that governmental entities are formulating pragmatic and actionable policies aimed at enhancing efficiency and sustainability within the construction sector. These initiatives stem from a recognition of BIM's potential in bolstering major construction project delivery. The widespread acceptance and endorsement of BIM are founded on its capacity to facilitate real-time management decisions in construction, encompassing the utilization of anticipated simulation methodologies for projects. The results underscore the significance of interoperability in BIM, thus mitigating the necessity for a unanimous consensus. Moreover, BIM appears as the core tool for creating a collaborative working environment in the construction sector for major project stakeholders. The strategy ascertains that construction works deliver quality projects while saving on financial resources.

TM indicates that BIM model governance is a critical factor for the success of BIM pervasive major project delivery. The model facilitates the identification of problems experienced during the project and provides communication tools for interaction with other interfaces. The approach ascertains that the management of BIM projects is executed effectively while considering the recent developments and changes in BIM technology. Furthermore, the project highlights how data mining can be embraced in advanced project management to support data exploration and communication for better optimization, prediction, and understanding of the physical operations for the BIM pervasive major project delivery.

In the context of the BIMKG, the experiment's results contribute to a nuanced understanding of dynamic relationships among BIM terms and their strategic implications. As a theoretical method, the BIMKG aligns seamlessly with the BIMBOK and reinforces the objectives of the BIMTAPE Framework. By comparing insights with the Periodic Table of BIM, the experiment reveals a comprehensive view of the interconnectedness of BIM elements. This understanding is crucial for the BIMTAPE Framework's goal of enhancing performance in BIM pervasive major project delivery. The BIMKG's ability to visualize relationships supports decision-making and knowledge engineering, offering a dynamic tool for navigating the intricate landscape of BIM. The integration of BIMKG insights into the framework equips practitioners with a strategic perspective, fostering continual improvement in BIM understanding and implementation for more efficient major project delivery.

4.5 Experiment 2: BIM TS 2 Technical Files from Case Projects

4.5.1 Introduction

A large part of BIMTAPE Framework, and the main concern of this chapter, involves an empirical analysis of the technical files from case projects, which are digital content of Crossrail website, with regard to performance enhancement for BIM pervasive major project delivery. Although these content data are freely available online, access to the data requires web scraping techniques and computational power. The data was obtained through web scraping techniques using Python and RStudio due to the large size of the data, and limitations on free public access to the website. This analysis provides valuable insights into learning legacies from major projects and how they can be applied to improve BIM-oriented performance.

The experiment described the process of web scraping content and metadata, cleaning and analysing the large set of unstructured data using techniques in textual analysis and classification. The aim was to extract representations that explain the textual information, allowing for further analysis and insights.

The purpose of textual analysis is to convert natural language text into a computer-friendly data structure that describes its meaning. In ML research for linguistics, simpler representations of text are extracted and fed into the training model, including characterizations of syntactic information (such as part-of-speech tagging and chunking) and semantic information (such as word-sense disambiguation and name entity extraction). These representations help the model understand the meaning of the text (Collobert et al., 2011).

The training model uses statistical estimation to categorize and visualize data based on the linguistic variables extracted from the text. Textual categorization involves pattern recognition, where machines recognize common phrases or topics within a group of similar documents. NLP methods and document classification via ML techniques were applied in this experiment. The first layer of NLP converted documents into term frequency and topic representations, while the second layer of ML used an SVM to predict document labels based on the term frequency and topic probability from the first layer.

The purpose of the second classification in this experiment was to standardize document labels and predict missing or similar labels. NLP is the computational manipulation of natural languages, with two main approaches: rule-based and statistical. The focus in this experiment was on using statistical NLP, which provides advantages for unsupervised and semi-supervised learning. Python was used as the programming language for NLP and RStudio packages were also employed to complete some tasks. Publicly available code sources were used to carry out the experiment, as the researcher lacked coding experience.

This experiment provides a contribution to the current body of knowledge through developing a TM approach for BIM-integrated performance enhancement in major construction project delivery. The main aim of this experiment is to develop a TM approach to improve BIM-integrated performance in major construction project delivery. The study intends to identify business environments related to BIM and major construction project delivery that can assist in the analysis. The data for the experiment was collected from the “Learning Legacy” website of Crossrail, a case-based experience in the context of BIM pervasive major project delivery.

The following section explains the tools used for the experiment, then the web scraping techniques used to collect the data for the experiment are explained. This is followed by an outline of the text analysis process and text visualization techniques used, concluding with a summary of the lessons learned from the experiment.

4.5.2 Tools for Study

The experiment utilised the following software tools and platforms for bibliography management, data analysis, document preparation, and text processing:

(1) Bibliography Management Mendeley Desktop
(2) Data Analysis Python 3.7.4 and R
(3) Document Preparation Txt and MS Excel.
(4) Python Packages and Libraries re, Gensim, BeautifulSoup, spaCy, Numpy, NLTK, and Pandas.
(5) R packages and libraries

These tools are commonly used in various fields of research and data analysis, particularly in the fields of NLP, data visualization, and data manipulation. Mendeley Desktop is a tool for managing and organizing bibliographic references, making it easier for researchers to keep track of the sources they have used in their work. Python and RStudio are both popular programming languages for data analysis. and provide a wide range of libraries and packages for various tasks such as text processing, data manipulation, and visualization. MS Excel is a common software package for data manipulation and preparation, while text files and plain text editors are often used for document preparation and data storage. The listed Python and R packages and libraries are commonly used for specific tasks such as regular expression handling, topic modelling, web scraping, NLP, numerical computation, and data manipulation.

4.5.3 Body of Text

4.5.3.1 Data Collection

The first step in data collection of Crossrail content involves web scraping. In this section, the detailed process of obtaining article content and metadata from the Crossrail website is described. Crossrail is the largest railway construction project underway in the UK, mainly in central London. The Crossrail website gives third parties permissions to use services from the website without the need to register. This website was created in accordance with conventional web standards, utilizing correct

XHTML for the text, CSS for the presentation, and minimal and unobtrusive JavaScript for the behaviour of the site. In addition, they made the language on the website as easy as possible to read and comprehend.

Once the webpage was accessed, the researched used Python to acquire metadata, which was the first step in scraping the Crossrail website. Metadata include information about dates, titles, authors, document types, sections, URLs, etc., of published articles. In the second scraping step, the collected URLs to access the paper's individual articles and extract the texts were used. The step can be done locally from a home computer. To maximize time and efficiency, this step involves the transfer of a large amount of data that is used for TM purpose.

Since this thesis is concerned mainly with text analytics, only the body of text from each article was gathered. This step requires finding the URLs with regard to the use of BIM in major construction project delivery. Each case study that makes use of BIM has been granted its own website, which can be accessed by the URL. The search term used to retrieve these documents was 'building information modelling' OR 'BIM' OR 'building information management' (Figure 4.7). A total of six webpages, containing up 56 articles, were scraped from the website for the period from January 2014 to August 2022. The retrieved documents from deploying the search terms and strategy comprised eight case studies, five good practise documents, three journal publications, four micro-reports, and 36 technical papers, as presented in the previous chapter (Table 3.1). Each webpage of 56 documents was scraped individually using Python.

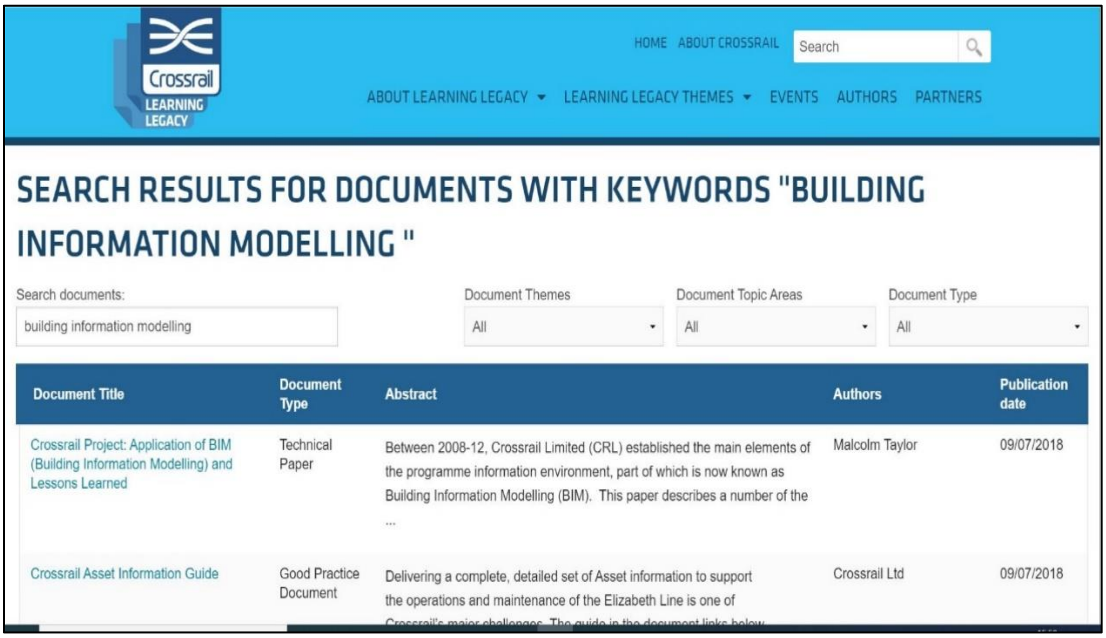


Figure 4.7: Data gathering from Crossrail (Crossrail Ltd., 2022)

The researcher began by importing standard Python libraries for reading and manipulating data, including Pandas, Numpy, and boilerpy3, etc. In order to request full texts of individual articles from Crossrail, the researcher firstly imported all the necessary libraries (see Figure 4.8), then each returned URL from the search terms was copied in Python in order to get the text body (see Figure 4.9).

```python
from boilerpy3 import extractors
import requests
from bs4 import BeautifulSoup
import pandas as pd
import numpy as np
from nltk.tokenize import word_tokenize
from wordcloud import WordCloud, STOPWORDS, ImageColorGenerator
import matplotlib.pyplot as plt
from collections import Counter
import seaborn as sns
from nltk.corpus import stopwords
import re
import string
import nltk
from sklearn.feature_extraction.text import CountVectorizer, TfidfVectorizer
import matplotlib.pyplot as plt
from wordcloud import WordCloud
import stylecloud
from urllib.request import urlopen
from sklearn.feature_extraction import text
import requests
from bs4 import BeautifulSoup
import pandas as pd
from pyquery import PyQuery as pq
import re
import matplotlib.pyplot as plt
%matplotlib inline
```

Figure 4.8: Standard Python libraries

```python
url1 = 'https://learninglegacy.crossrail.co.uk/documents/ground-settlement-behaviour-in-chalk-due-to-tbm-excavations/'
url2 = 'https://learninglegacy.crossrail.co.uk/documents/the-protection-of-the-400kv-cables-at-pudding-mill-lane/'
url3 = 'https://learninglegacy.crossrail.co.uk/documents/information-handover-principles/'
url4 = 'https://learninglegacy.crossrail.co.uk/documents/building-a-spatial-infrastructure-for-crossrail/'
url5 = 'https://learninglegacy.crossrail.co.uk/documents/vibration-management-and-listed-buildings/'
url6 = 'https://learninglegacy.crossrail.co.uk/documents/supply-chain-quality-requirements/'
url7 = 'https://learninglegacy.crossrail.co.uk/documents/the-importance-of-construction-mock-ups-and-trials/'
url8 = 'https://learninglegacy.crossrail.co.uk/documents/data-management-analysis-and-visualisation-on-crossrail-drive-x-western-
url9 = 'https://learninglegacy.crossrail.co.uk/documents/design-of-crossrails-precast-tunnel-linings-for-fire/'
url10 = 'https://learninglegacy.crossrail.co.uk/documents/design-sprayed-concrete-linings-soft-ground%e2%94%80-crossrail-perspect
url11 = 'https://learninglegacy.crossrail.co.uk/documents/stepney-green-cavern-design-concepts-and-performance-of-scl-lining/'
url12 = 'https://learninglegacy.crossrail.co.uk/documents/heritage-deed/'
url13 = 'https://learninglegacy.crossrail.co.uk/documents/heritage-deed-for-settlement-mitigation-works-affecting-listed-building
url14 = 'https://learninglegacy.crossrail.co.uk/documents/design-crossrail-farringdon-station-engineers-perspective-2/'
url15 = 'https://learninglegacy.crossrail.co.uk/documents/young-crossrail-programme/'
url16 = 'https://learninglegacy.crossrail.co.uk/documents/addressing-skills-gaps-direct-intervention-tuca/'
url17 = 'https://learninglegacy.crossrail.co.uk/documents/noise-vibration-controls-tbm-temporary-construction-railway/'
url18 = 'https://learninglegacy.crossrail.co.uk/documents/lindsey-street-bridge-structural-solution-settlement-mitigation/'
url19 = 'https://learninglegacy.crossrail.co.uk/documents/design-construction-fisher-street-crossover-cavern-crossrail-contract-c
url20 = 'https://learninglegacy.crossrail.co.uk/documents/design-deep-cut-cover-crossrail-paddington-station-using-finite-element
url21 = 'https://learninglegacy.crossrail.co.uk/documents/3d-geological-model-completed-farringdon-underground-railway-station/'
url22 = 'https://learninglegacy.crossrail.co.uk/documents/covered-way-126-safeguarding-brittle-structure-live-railway-ground-move
url23 = 'https://learninglegacy.crossrail.co.uk/documents/innovative-verification-process-speeds-construction-crossrails-moorgate
url24 = 'https://learninglegacy.crossrail.co.uk/documents/bim-metrics/'
url25 = 'https://learninglegacy.crossrail.co.uk/documents/crossrail-bim-principles/'
url26 = 'https://learninglegacy.crossrail.co.uk/documents/crossrail-development-bim-environment/'
url27 = 'https://learninglegacy.crossrail.co.uk/documents/crossrail-lfb-emergency-services-liaison/'
```

Figure 4.9: Returned URLs from a search term

The researcher then extracted all articles from the links. Boilerpy3 package was used for this purpose by using extractors. The “get_content ()” methods were used to extract the filtered text. For example, the following script (see Figure 4.10) creates an “ArticleExtractor” instance, which is tuned towards BIM articles. This extraction strategy works very well for most types of article-like HTML documents (Riebold, 2021). The used code iterated through the list and appended the elements to the string. Now that the researcher has written the code to open the URL, it is time to extract the data from the website. The researcher used the Boilerpy3 package to extract the content of articles from a list of URLs. The “ArticleExtractor” instance was created and tuned specifically for BIM articles, and the “get_content()” method was used to extract the filtered text. The code then iterated through the list of URLs and appended the extracted elements to a string. This is a common way to extract data from websites, and the Boilerpy3 package can be a useful tool for this purpose.

```
extractor = extractors.ArticleExtractor()
response = requests.get(url1)
content1 = extractor.get_content(response.text)
main_content_string =  "\n\n".join(content1)
main_content_string = re.sub('\s+',' ',str(content1)).strip()
main_content_string[:]
```

```
'Ground Settlement Behaviour In Chalk Due To TBM Excavations Document type: Technical Paper Abstract In Crossrail project con
tract C310, two new tunnels - the Thames Tunnel (westbound and eastbound) will be constructed between Plumstead portal and No
rth Woolwich portal. The Thames Tunnel is 2.6 km long and about 15 m below the existing Thames river bed. Construction of Plu
mstead portal commenced in 2011 and completed in 2012 for the westbound TBM launching in January 2013 and eastbound TBM launc
hing in May 2013. The construction of North Woolwich portal has begun in early 2012 in order to receive the first TBM in 201
4. The running tunnels will be constructed predominantly through chalk. Two slurry TBM machines, Sophia and Mary named after
the wives of Marc Isambard Brunel and Isambard Kingdom Brunel will be adopted for the sake of the existing ground condition -
majority in chalk. The degree of settlement control is one of the prerequisites for tunnelling design for the consequent risk
of damage to third-party assets. To obtain the surface and sub-surface settlement profile and corresponding volume loss due t
```

Figure 4.10: Script for article extraction

As mentioned earlier, the desired data to extract is text only. Now that a list of documents was gathered, it is time to create Panda’s data frames, and save them as csv files. Panda’s data frames are powerful and versatile for storing and working with data in Python. This function takes the input values of the article's title, abstract, document type, full text, author, and the link or path of the article. The output value is a data frame with columns indicating the variables in the article metadata and rows indicating the BIM articles. Also, the extracted data was stored in .txt format (Figure 4.11), which was achieved by copying the outcomes from Python and pasting them in notebook. Each URL has its own notebook.

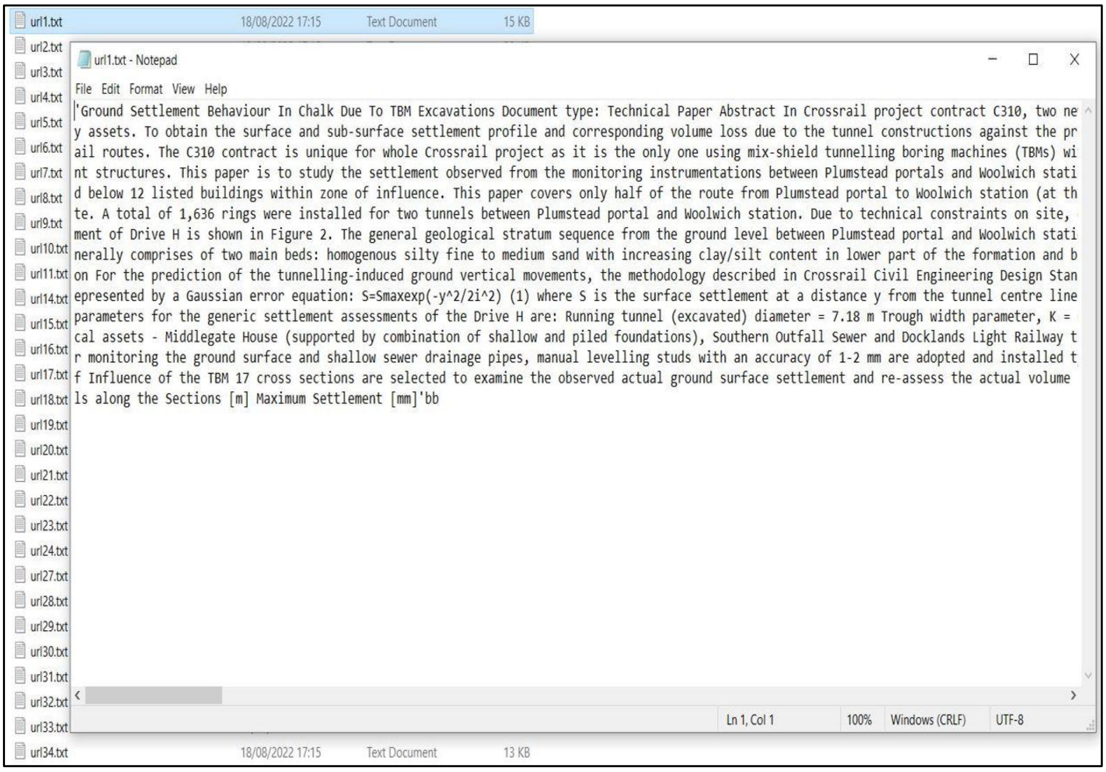


Figure 4.11: Output value of article metadata

4.5.3.2 Merge Seven Files

However, some URLs have external documents inside, and articles within these URLs cannot be scraped automatically. Seven articles in pdf format were found. Therefore, to ensure impartiality, the researcher manually retrieved these seven articles in PDF format and transferred their content into individual .txt files. These files were then combined into one using a Python code (Figure 4.12). Despite this, the merged file had too much unnecessary information and needed to be purified before any text analysis could take place. The final purified dataset can be accessed via Mendeley using the provided link at https://data.mendeley.com/datasets/5hcg7tddc4/1.

```
import glob

dir = r'C:\Users\Ambark Bareka\driver\Sevenmerged'

files = glob.glob(dir + '\\*.txt')

file_big = 'sevenmerged.txt'

with open(file_big, 'wb') as fnew:
    for f in files:
        with open(f, 'rb') as fold:
            for line in fold:
                fnew.write(line)
                fnew.write("\n".encode(encoding='utf_8'))
```

Figure 4.12: Combination code

In this phase of the research, data was meticulously gathered from the Crossrail website through the use of web scraping techniques. In the next stage, TM analysis will be performed on the obtained data, which consists of 56 documents represented by 56 URLs. To carry out the TM analysis, appropriate Python libraries were imported. The TM software possesses an analyser function that is capable of processing any text document in a specific format, such as plain text files. The Crossrail articles dataset was then consolidated into a single variable called “raw text”. To guarantee that all necessary information was acquired accurately, the data type, length, and the first 1000 characters of raw text will be thoroughly scrutinized.

4.5.4 Process of TM

In this phase of the research, the researcher utilized NLP and ML techniques to conduct a thorough analysis of the textual data that was previously collected. This requires high computational power, thus Python Jupyter Notebook (Anaconda) and RStudio were used whenever possible to extract meaningful insights from the scraped data in a systematic and automated manner. The dataset was uploaded to the Python workspace, and the researcher proceeded to clean, analyse, and visualize the textual data. The cleaning process will start by removing any unwanted characters from the Crossrail texts using the Regular Expression (Aho, 1991) module in Python. This includes any embedded URLs that are present in the body of the texts. Once the cleaning is complete, the researcher will proceed to analyse the texts through Natural Language Processing (NLP).

4.5.4.1 Natural Language Processing (NLP)

The researcher began by processing raw data, consisting of the written content of Crossrail-BIM articles. The researcher compiled all articles from the website into a list of texts, each of which represents a separate article. There are a total of 56 articles about the implementation of BIM in large-scale project delivery from Crossrail. The researcher utilized the spaCy (Honnibal and Montani, 2017), Gensim (Rehurek and Sojka, 2011), and scikit-learn (Pedregosa et al., 2011) libraries to analyse this list of texts. Firstly, the researcher loaded the English tokenizer, tagger, parser, NER, and word vectors that come with the spaCy library, as described in the previous section and depicted in Figure 4.10.

Afterwards, the researcher eliminated the line breaks within each article and saved the documents to disk, with each article separated by a line break. This resulted in a large text file, wherein each line corresponds to a separate article. The revised texts were thus prepared for NLP analysis. The researcher employed the basic text processing functions from the Text-Analytics-Module for Python (Wang, 2018) for the purpose of textual analysis and topic visualization. The module was imported into Python, and the configurations for NLP were redefined.

The corpus was then thoroughly cleaned for analysis via several steps, including removal of punctuation, stopwords, normalization of the corpus, substitution of entities with tokens, parsing of documents, lemmatization of texts, and extraction of sentences. The output was used to generate unigrams of sentences, where a unigram is a single syllable sequence from a text corpus in computational linguistics. Additionally, the retrieved text contained numerous special characters that were not necessary for preservation and would only complicate the analysis. These characters, which included \n and \r for formatting and [and] for displaying metadata were all removed. The raw text also contained information such as the document type, author, link, abstract, and date, which were not needed for the analysis and were therefore deleted as well.

The code for this process is shown in Figure 4.13. After the removal of special characters, the first 1000 characters were re-examined to observe the changes to the data. At this point, it can be stated that the researcher has completed the cleaning process, and the Crossrail BIM data is now free of any special characters or unusual spacing.

```
# Replace the special characters with a space

raw_text = re.sub("\\n|\\r|\[.+?\]", " ", raw_text)

# Replace any set of two or more spaces with just a single space

raw_text = re.sub(" {2,}", " ", raw_text)
```

```
# Re-examine the first 1000 characters

raw_text[:1000]
```

```
'Ground Settlement Behaviour In Chalk Due To TBM Excavations Document type: Technical Paper Abstract In Crossrail project contr
act C310, two new tunnels - the Thames Tunnel (westbound and eastbound) will be constructed between Plumstead portal and North
Woolwich portal. The Thames Tunnel is 2.6 km long and about 15 m below the existing Thames river bed. Construction of Plumstead
portal commenced in 2011 and completed in 2012 for the westbound TBM launching in January 2013 and eastbound TBM launching in M
ay 2013. The construction of North Woolwich portal has begun in early 2012 in order to receive the first TBM in 2014. The runni
ng tunnels will be constructed predominantly through chalk. Two slurry TBM machines, Sophia and Mary named after the wives of M
arc Isambard Brunel and Isambard Kingdom Brunel will be adopted for the sake of the existing ground condition - majority in cha
lk. The degree of settlement control is one of the prerequisites for tunnelling design for the consequent risk '
```

Figure 4.13: Code for cleaning cross rail BIM data

However, unigrams on their own do not fully convey the conceptual meaning of sentences, and it is necessary to examine groups of words that stand together as a conceptual unit, such as a phrase. The idea behind phrase detection is to determine if a group of tokens constitutes a meaningful phrase. To accomplish this, the researcher will use phrase modelling techniques (Harrison, 2016) by iterating through the tokens in the text corpus and identifying words that co-occur together more frequently than would be expected by chance.

The formula for the phrase model to determine if two tokens A and B form a phrase is as follows:

$$\frac{count(A\,B) - count_{min}}{count(A) * count(B)} * N > threshold$$

Where:

- $count(A)$ is the number of times token A appearing in the corpus.
- $count(B)$ is the number of times token B appearing in the corpus.
- $count(A\,B)$ is the number of times the tokens A B appear in the corpus in order.
- N is the total size of the corpus vocabulary.
- $count_{min}$ is a user-defined parameter to ensure that accepted phrases occur a minimum number of times.
- $threshold$ is a user-defined parameter to control how strong of a relationship between two tokens the model requires before accepting them as a phrase.

The Gensim library offers tools for phrase modelling, which are trained on a text corpus (Rehurek and Sojka, 2011). The model is then applied to new text. If it identifies two tokens in the new text as a phrase, it will merge the two tokens into a single new word. The creation of the phrase model is achieved by executing the Python code shown in Figure 4.14. The output is a text that is lemmatized, has phrases, and stop-words removed from the original text.

```
# Create a keywords column that only contains significant words

# Split the text into words

BIM_df["keywords"] = BIM_df["clean_sentence"].apply(word_tokenize)

# Remove insignificant words ("stopwords") and one-letter words

stop = stopwords.words("english")  # nltk stopwors list
stop.extend("acknowledgement")  # Add specific words

BIM_df["keywords"] = BIM_df["keywords"].apply(lambda x: [w for w in x if w not in stop and len(w) > 2])
```

Figure 4.14: Create a keywords column that only contains significant words

The next step is to construct a dictionary for the lemmatized, phrased, and stop-word removed text corpus, followed by creating a serialized corpus. The code that follows lemmatizes the words and generates a set of keywords for each article. Upon examining the results, it can be seen that the data frame includes the original raw sentence, cleaned sentence, and a list of keywords for each article (as seen in Figures 4.15 and 4.16). The researcher then created a new column with the number of keywords per sentence (as seen in Figure 4.17).

```
# Lemmatize the words, reducing them back to their stems

# Object to lemmatize words

lemma = WordNetLemmatizer()

# Dictionary to convert between part-of-speech tags and the ones the lemmatizer understands

tag_dict = defaultdict(lambda : wordnet.NOUN)
tag_dict['J'] = wordnet.ADJ
tag_dict['V'] = wordnet.VERB
tag_dict['R'] = wordnet.ADV

# function to lemmatize a sentence

def get_lemma(word):
    """Gets the POS tag for a word, and then returns the lemmatized form of the word"""
    tag = pos_tag([word])[0][1][0]
    tag = tag_dict[tag]

    return lemma.lemmatize(word, tag)

# Actually perform the lemmatization

BIM_df["keywords"] = BIM_df["keywords"].apply(lambda x: [get_lemma(word) for word in x])
```

Figure 4.15: Apply lemmatization to a corpora

```
# Inspect the keywords column

BIM_df.sample(10)
```

	raw_sentence	articl	date	clean_sentence	keywords
1914	crossrail found that there was a lack of experience in the market and therefore decided to open negotiations with construction industry training board citb to take over the running of tuca, with support from crossrail.	10		crossrail found that there was a lack of experience in the market and therefore decided to open negotiations with construction industry training board citb to take over the running of tuca with support from crossrail	[crossrail, found, lack, experience, market, therefore, decide, open, negotiation, construction, industry, training, board, citb, take, run, tuca, support, crossrail]
2453	to mitigate this risk, bfk and dal developed a detailed probing ahead and de pressurisation strategy to reduce the residual risk as low as reasonably practicable alarp .	13		to mitigate this risk bfk and dal developed a detailed probing ahead and de pressurisation strategy to reduce the residual risk as low as reasonably practicable alarp	[mitigate, risk, bfk, dal, developed, detailed, probe, ahead, pressurisation, strategy, reduce, residual, risk, low, reasonably, practicable, alarp]
2449	the original concept also included the provision for compensation grouting within the london clay above the crossover.	13		the original concept also included the provision for compensation grouting within the london clay above the crossover	[original, concept, also, include, provision, compensation, grout, within, london, clay, crossover]
4867	mechanisms can be built in to the model diagrams to aid navigation across the various levels.	28		mechanisms can be built in to the model diagrams to aid navigation across the various levels	[mechanism, built, model, diagram, aid, navigation, across, various, level]
7169	sorba and concrete valley took an innovative approach to this problem.	40		sorba and concrete valley took an innovative approach to this problem	[sorba, concrete, valley, take, innovative, approach, problem]
7754	dpf consumer electricity networks which have distorting and displacing load currents, will have a reduced true power factor.	43		dpf consumer electricity networks which have distorting and displacing load currents will have a reduced true power factor	[dpf, consumer, electricity, network, distort, displace, load, current, reduce, true, power, factor]
3559	using common design methods would improve efficiency and reduce work load, thus reducing cost.	21		using common design methods would improve efficiency and reduce work load thus reducing cost	[use, common, design, method, would, improve, efficiency, reduce, work, load, thus, reduce, cost]
2108	where such events were audible, this will be mainly due to the tonal characteristics, as the measured overall a weighted noise level are comparable to guideline internal noise levels recommended by who and in bs8233.	11		where such events were audible this will be mainly due to the tonal characteristics as the measured overall a weighted noise level are comparable to guideline internal noise levels recommended by who and in bs	[event, audible, mainly, due, tonal, characteristic, measure, overall, weight, noise, level, comparable, guideline, internal, noise, level, recommend]
5354	although surface ejector wells were able to lower the groundwater level below eastbound tunnel inverts, it was not sufficient for the westbound tunnel which is situated 5m deeper.	31		although surface ejector wells were able to lower the groundwater level below eastbound tunnel inverts it was not sufficient for the westbound tunnel which is situated m deeper	[although, surface, ejector, well, able, low, groundwater, level, eastbound, tunnel, inverts, sufficient, westbound, tunnel, situate, deeper]

Figure 4.16: Inspecting the outcome of keyword's column

```
# Create a new column containing the number of keywords per sentence

BIM_df["keyword_count"] = BIM_df["keywords"].apply(lambda x: len(x))
```

Figure 4.17: Code for new column data

The researcher then proceeded to eliminate rows with no keywords, rows that simply contain the name of a month, and filtered the dataset to retain only the columns where the keywords column is not a subset of months (as seen in Figure 4.18).

```python
# Drop rows with zero keywords

BIM_df = BIM_df[BIM_df["keyword_count"] > 0]

# Drop rows that simply contain the name of a month

# Set holding months

months = {"january", "february", "march", "april", "may", "june",
          "july", "august", "september", "november", "december"}

# Filter the dataset, keeping only the columns where the keywords column
# is not a subset of months

BIM_df = BIM_df[~BIM_df["keywords"].map(lambda x: set(x).issubset(months))]
```

Figure 4.18: Code for data categorization

4.5.4.2 Creating Dictionary

The researcher then created a dictionary of words from the cleaned text corpus (as seen in Figure 4.19). This step helped identify all the unique words present in the text. Using this dictionary, the cleaned text in each row was then transformed into a Document-Term Matrix (DTM), also known as a Bag of Words (BoW), which essentially stores the frequency and occurrence of each word in the corpus. To find the word frequencies, the keywords column was transformed into a document-term matrix, and an object that converted strings into a sparse matrix was also created. The keywords series was then converted into a sparse matrix (as seen in Figure 4.20). Finally, a data frame containing the terms and frequencies was generated.

```python
# Convert the keywords column into a document-term matrix

keywords_series = BIM_df["keywords"].apply(lambda x: " ".join(x))

# Create an object that converts strings into a sparse matrix

vectorizer = CountVectorizer(max_features= 10000 , min_df=0.005, ngram_range= (1,1))

# Convert the keywords series to a sparse matrix

BIM_dtm = vectorizer.fit_transform(keywords_series)
```

Figure 4.19: Code for dictionary of words

```python
# Create a dataframe of terms and frequencies

# Get the frequency of each term

keyword_frequencies = BIM_dtm.sum(axis=0)

# Match frequencies to terms

frequent_terms = [(word, keyword_frequencies[0, id]) for word, id in vectorizer.vocabulary_.items()]

# Convert the word-frequency pairings to a dataframe

keyword_df = pd.DataFrame(columns=["term",
                                   "frequency"], data=frequent_terms)

# Sort the dataframe

keyword_df.sort_values(by="frequency", ascending=False, inplace=True)
```

Figure 4.20: Code for bag of words

4.5.4.3 Generating Top 20 Frequent Words

The results of the TM analysis illustrated in Table 4.15 appear to be disappointing, as the first specific word, “design”, is over twenty places down in the list. This may suggest that the use of BIM in major construction project delivery focuses more on the design stage rather than performance-related issues. The presence of terms such as “crossrail” and “figure” in the analysis also indicate that these terms should be added to the list of stopwords to be excluded from future analyses, as they do not appear to be relevant to the study. In light of these findings, the researcher suggests that searching directly for words related to performance enhancement such as cost, time, and quality (Chen, 2022) may yield more meaningful results in the TM approach for performance enhancement in BIM pervasive major project delivery (see Figure 4.21).

Table 4.15: Frequency of top 20 used words

No.	Words	Frequencies	No.	Words	Frequencies
1	Design	1285	11	Station	641
2	System	1230	12	Room	629
3	Tunnel	1144	13	Information	599
4	Use	1098	14	Data	580
5	Crossrail	1061	15	Provide	548
6	Figure	839	16	Level	521
7	Model	760	17	Require	477
8	Project	754	18	Ground	467
9	Construction	705	19	Line	454
10	Work	684	20	Asset	453

```
# Create a list of plot-specific keywords

plot_keywords = ["cost", "time", "quality", "enhancement", "bim", "performance", "efficiency", "visualisation", 'Man hours',
                 'Development', 'product', 'process', 'people', 'Client satisfaction', 'speed', 'corrections', 'errors',
                 'sharing', 'flexibility', 'financial', 'disciplines','CAD', 'satisfaction', 'bids']


# Return frequencies for selected terms

keyword_df[keyword_df["term"].isin(plot_keywords)]
```

Figure 4.21: Code for searching specific words

The text analysis shown in Table 4.16 provides more insights and relevance to the study. The word "process" appeared the most frequently with 405 appearances, followed by "time" with 311 appearances, which may indicate that the delivery of major projects is often delayed due to poor performance. This is further supported by the frequent appearance of "performance" (251 times) and the less frequent appearance of "product" (63 times), "efficiency" (57 times), and "people" (41 times). "Quality" (185 times) and "cost" (180 times) were also found to be important performance-related issues within the context of major construction project delivery. The appearance of "BIM" (124 times) suggests that although the technology has been adopted and utilized, there are still challenges in improving efficiency.

Table 4.16: Results of specific searched words

No	Words	Frequencies
1	Process	405
2	Time	311
3	Performance	251
4	Quality	185
5	Cost	180
6	BIM	124
7	Product	63
8	Efficiency	57
9	People	41

4.5.5 New Knowledge

4.5.5.1 Topic Modelling

The Latent Dirichlet Allocation (LDA) model is a commonly used ML technique for topic modelling, where it helps to identify underlying themes or topics in a collection of documents. The LDA model works by taking the observed words in documents and explaining each document as a mixture of multiple topics. Each topic is then represented by a number of representative words. In this experiment, the researcher applied the LDA model to the textual corpus to construct the topics for the data. The training of the LDA model was conducted by running the appropriate lines of code in a Jupyter Notebook, as shown in Figure 4.22.

```
topic_dict = corpora.Dictionary(lda_docs)
# Use the dictionary to generate a corpus from the texts
# Convert the keywords into dictioanary references
topic_corpus = [topic_dict.doc2bow(text) for text in lda_docs]
numtopics = 30
ldafile = 'topic_corpus' +str(numtopics)
lda = LdaModel(topic_corpus, num_topics=numtopics, id2word=topic_dict, passes=100)
lda.save(ldafile)
lda.minimum_probability = 0.0
```

Figure 4.22: Command for LDA model

The researcher used the Document-Term Matrix (DTM) to mine the top 10 terms related to each topic and classify the articles' expressive topics. However, the initial results showed that the keywords did not make any sense, with words such as “London”, “review”, and “crossrail” topping the list (Tables 4.17 and 4.18). To improve the results, the researcher decided to delete these words from the corpus and re-run the analysis (Table 4.19). This led to the identification of five main research themes: tunnel, ground, settlement, excavation, and construction; tunnels, monitoring, data, works, and caption (Topic_0); design, construction, lining, vibration, tunnel, and station (Topic_1); asset, data, and product (Topic_2); room, air (Topic_3); and design (Topic_4). However, despite the LDA model identifying each topic, the output still requires human input to label each topic. This can be done by two researchers individually labelling the topics based on their judgment. The ML approach provides a comprehensive picture of the research area and helps stakeholders and participants understand new development themes and hidden structures. It also gives an assessment of the area researched (Antons and Breidbach, 2018; Mahr et al., 2019; Gallouj and Savona, 2009).

Table 4.17: Returned RStudio topics before deletion

No.	**Topic 0**	**Freq.**		**Topic 1**	**Freq.**		**Topic 2**	**Freq.**
1	London	0.065		Slab	0.048		Review	0.049
2	New	0.055		Construction	0.043		Order	0.036
3	Building	0.035		Level	0.030		Follow	0.034
4	Tunnel	0.026		Support	0.027		Term	0.030
5	Central	0.025		Work	0.026		Long	0.027
6	Crossrail	0.024		Temporary	0.025		Require	0.026
7	Underground	0.021		Structure	0.023		Evidence	0.024
8	East	0.020		Construct	0.022		Product	0.024
9	Station	0.020		Platform	0.022		Large	0.023
10	Area	0.020		Prop	0.019		Day	0.023
No.	**Topic 3**	**Freq.**		**Topic 4**	**Freq.**		**Topic 5**	**Freq.**
1	Maintenance	0.068		Concrete	0.065		Project	0.050
2	Make	0.032		Line	0.058		Crossrail	0.049
3	Access	0.032		Method	0.045		Benefit	0.026
4	Array	0.028		Sprayed	0.027		Mechanical	0.025
5	Planning	0.024		Maximum	0.026		BIM	0.025
6	Fix	0.023		Use	0.022		Paper	0.024
7	Decision	0.022		Curve	0.022		Many	0.020
8	Also	0.020		Calculation	0.019		Programme	0.020
9	Aims	0.019		Layer	0.019		Learn	0.019
10	Addition	0.017		Value	0.018		Year	0.017

Table 4.17: Returned RStudio topics before deletion

No.	**Topic 6**	**Freq.**		**Topic 7**	**Freq.**		**Topic 8**	**Freq.**
1	Level	0.087		Life	0.041		Figure	0.153
2	Vibration	0.051		Step	0.037		Show	0.072
3	Noise	0.046		Whole	0.032		See	0.033
4	Failure	0.045		Length	0.028		Model	0.032
5	Performance	0.035		Survey	0.028		Use	0.025
6	Criterion	0.031		Along	0.025		Monitoring	0.023
7	High	0.027		Cycle	0.023		Green	0.022
8	Growth	0.027		Advance	0.022		Data	0.022
9	Table	0.020		Distribution	0.021		Stepney	0.021
10	Measure	0.019		Software	0.020		Section	0.018
No.	**Topic 9**	**Freq.**						
1	One	0.046						
2	Number	0.035						
3	Panel	0.029						
4	Improvement	0.025						
5	Occur	0.025						
6	Per	0.024						
7	Initiative	0.022						
8	Switch	0.021						
9	Individual	0.020						
10	Hour	0.020						

Table 4.18: Returned RStudio topics after deletion

No.	**Topic 0**	**Topic 1**	**Topic 2**	**Topic 3**	**Topic 4**
1	Tunnel	Design	Asset	Room	Design
2	Ground	Construction	Data	Air	Construction
3	Settlement	Lining	Product	System	Process
4	Excavation	Vibration	Document	Temperature	System
5	Construction	Tunnel	Management	Fire	Programme
6	Tunnels	Station	Time	Process	System
7	Monitoring	Tunnels	Power	Requirements	Layer
8	Data	Track	Assets	Rooms	Quality
9	Works	Ground	Quality	Equipment	Performance
10	Caption	Wall	Legacy	Heat	Team

4.5.5.2 Visualisation of Topic Modelling

The topics were represented by the most prominent words. To enhance comprehension, a visual representation of these topics was created following the approach of Chuang et al. (2012) and Sievert and Shirley (2014). The code used for this purpose is shown in Figure 4.23.

```
# Visualise topics
# This is a slow and computationally expensive process - my machine's fan goes crazy during it.
# Create the display

import pyLDAvis.gensim_models

lda_display = pyLDAvis.gensim_models.prepare(lda, topic_corpus, topic_dict, sort_topics=False)

# Actually display the visualisation

pyLDAvis.display(lda_display)
```

Figure 4.23: Visualization code for generated topics

Figure 4.24 displays an illustration of the visualization generated by the LDA model for one topic. The diagram is split into two parts: the left side depicts the proximity of the topics to each other, and the right side shows the 30 most relevant words that make up the topic. From the words in the topic, it is apparent that the topic pertains to different stages of construction, including design, pre-construction, procurement, construction, commissioning, owner occupancy, and project close-out. The presence of the word “performance” implies that BIM can be utilized for performance improvement.

The size of the circles in the visualization represents the proportion of the texts each topic covers, with topic 1 being the most dominant. The value of λ reflects the significance of the terms, with a λ value close to zero indicating that the terms are unique to the topic, while a λ value close to one means the terms appear in the topic with high probabilities. For a more balanced representation of the terms’ relevance and distinctiveness, a λ value of 0.6 is utilized. Some topics, like the one in Figure 4.25, may be straightforward for human interpretation, while others may not be as easily recognizable.

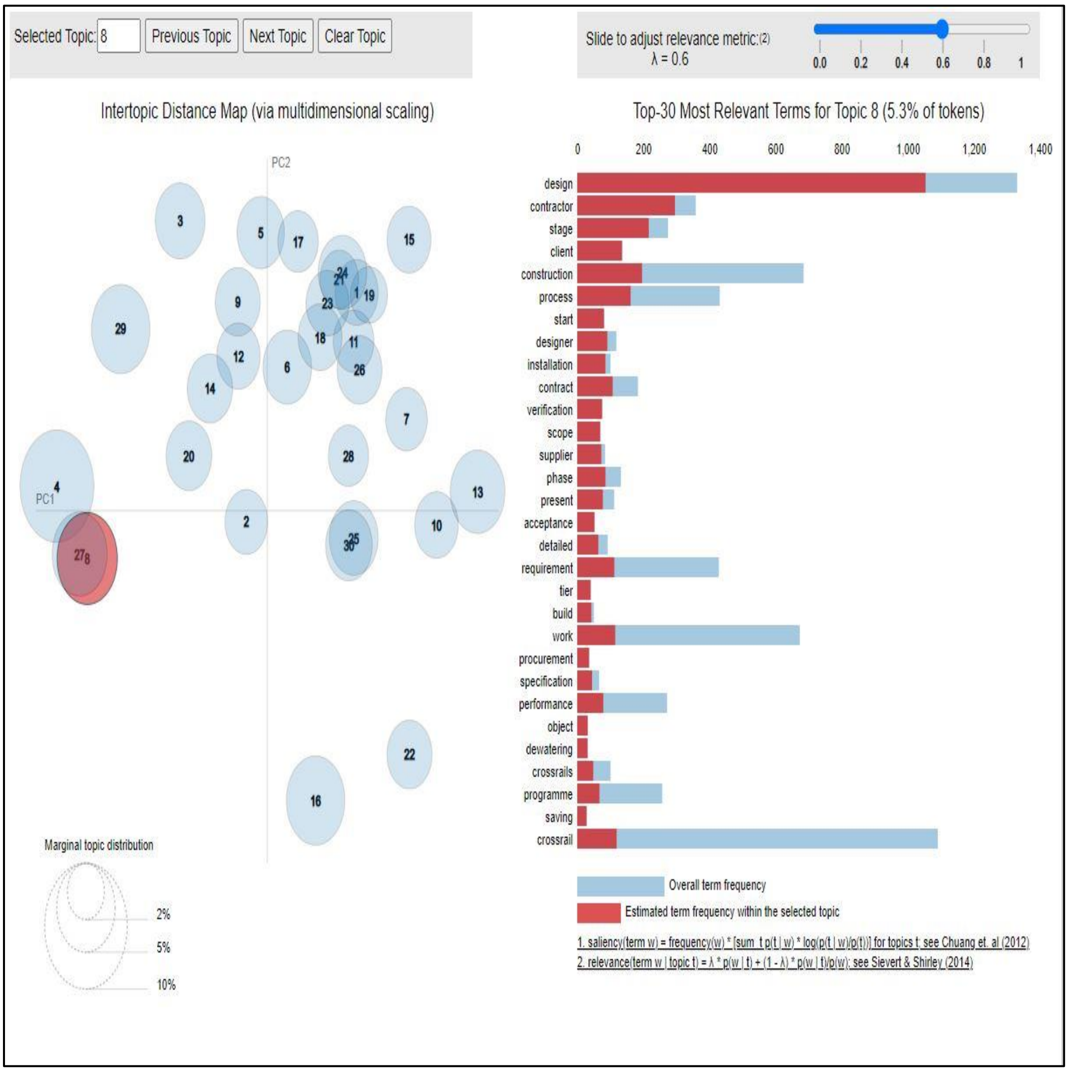


Figure 4.24: Visualization for topics created by the LDA model

4.5.5.3 Knowledge graph

In the realm of Experiment 2, the exploration involved employing topic modelling techniques to decipher prevalent themes within the corpus of BIM literature. The analysis revealed five distinctive topics, each characterized by a set of associated terms, encompassing aspects such as tunnel construction, design processes, asset management, room systems, and construction methodologies.

The resulting knowledge graph Figure 4.25, which was generated by the author using Python, constructed by establishing connections between these topics and common terms, visually represents the intricate relationships and influence among different facets of BIM. For example, the edge between the Tunnels node and the Ground node indicates that there is a relationship between these two topics. This graph aids in comprehending the complex interplay between BIM key terms, aligning with the objectives of the BIMTAPE Framework, which aims to enhance performance in BIM pervasive major project delivery.

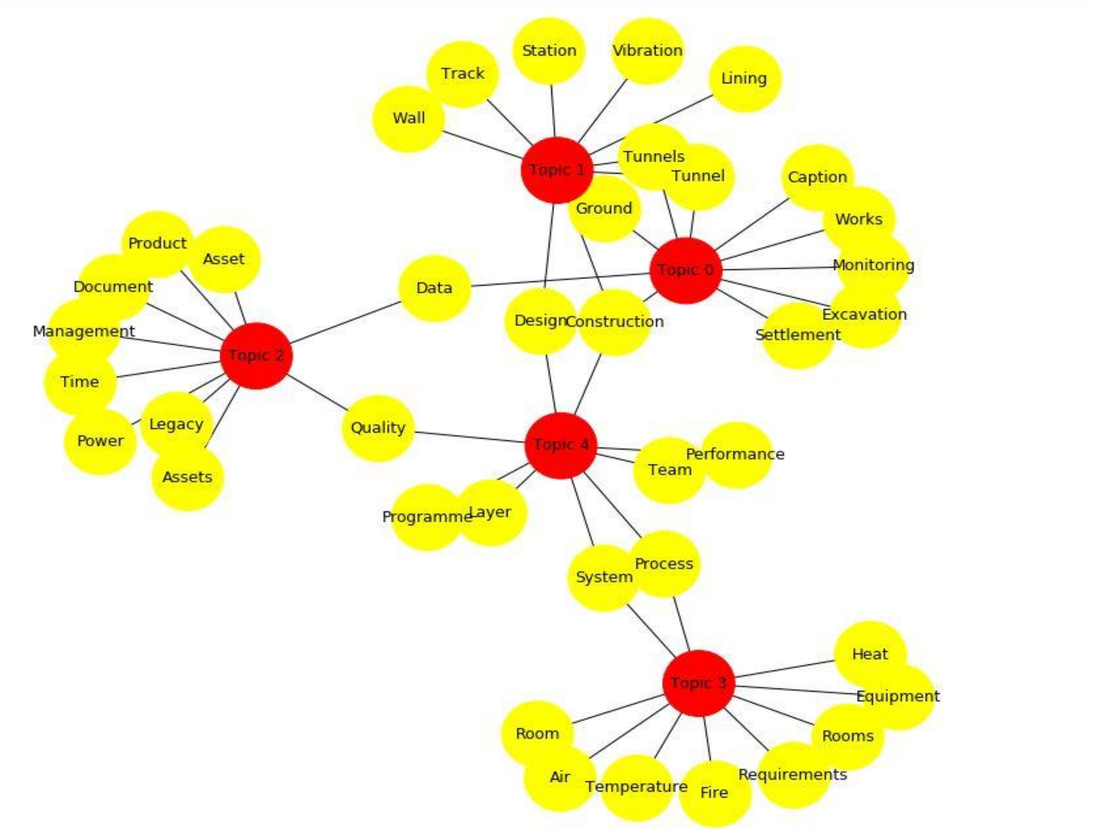


Figure 4.25: Connections between these topics and common BIM terms

Furthermore, a broader examination involves correlating the findings from the four distinct topic modelling tables (Topic 0 to Topic 4) with the key technical domains outlined in the Periodic Table of BIM (Mordue, 2016). This analysis provides valuable insights into the alignment and differences between the topic modelling results and the established BIM technical domains, enhancing the understanding of the nuanced relationships within the BIM landscape. Additionally, the Tables 4.19, 4.20, 4.21, 4.22, and 4.23 present the key technical terms from each topic along with their connections to the technical domains in the Periodic Table of BIM, further elucidating the intricate interplay between topic modelling outcomes and established BIMTAPE Framework.

The presented tables underscore the alignment and interplay between key technical terms derived from topic modelling and their corresponding technical domains outlined in the Periodic Table of BIM. Notably, terms such as Tunnel, Ground, and Settlement prominently align with the Foundation and Collaboration domains, emphasizing the foundational and collaborative aspects in the BIM landscape. The dynamic relationships captured by the BIMKG further enrich understanding, reinforcing the goals of the BIMTAPE Framework for performance enhancement in major project delivery.

The integration of TM results in these tables strategically links the BIMTAPE Framework with practical applications summarized by the Periodic Table of BIM (Mordue, 2016). By elucidating the nuanced relationships between TM-derived key terms and their alignment with technical domains, the BIMKG becomes a powerful bridge between theoretical frameworks and real-world BIM practice. This synergy amplifies the utility of the Periodic Table of BIM, rendering it a more robust and informed guide for practitioners and decision-makers, contributing significantly to the ongoing development of the BIMBOK.

Table 4.19: Results from Topic 0 compared to the Periodic Table of BIM

Priority	**Key Technical Terms from Topic Modelling**	**Technical Domains in the Periodic Table of BIM**
1	Tunnel	Foundation, Collaboration
2	Ground	Foundation, Collaboration, Process
3	Settlement	Foundation
4	Excavation	Foundation, Collaboration
5	Construction	Foundation, Collaboration, Process
6	Tunnels	Foundation, Collaboration
7	Monitoring	Foundation, Collaboration
8	Data	Foundation, Collaboration
9	Works	Foundation, Collaboration
10	Caption	Foundation

Table 4.20: Results from Topic 1 compared to the Periodic Table of BIM

Priority	Key Technical Terms from Topic Modelling	Technical Domains in the Periodic Table of BIM
1	Design	Foundation, Collaboration, Process
2	Construction	Foundation, Collaboration, Process
3	Lining	Foundation, Collaboration
4	Vibration	Foundation, Collaboration
5	Tunnel	Foundation, Collaboration
6	Station	Foundation, Collaboration
7	Tunnels	Foundation, Collaboration
8	Track	Foundation, Collaboration
9	Ground	Foundation, Collaboration
10	Wall	Foundation, Collaboration

Table 4.21: Results from Topic 2 compared to the Periodic Table of BIM

Priority	Key Technical Terms from Topic Modelling	Technical Domains in the Periodic Table of BIM
1	Asset	Process
2	Product	Process
3	Management	Process, People
4	Document	Process
5	Management	Process
6	Time	Process, Digital Plan of Work stages
7	Power	Process
8	Assets	Process
9	Quality	Process, Technology
10	Quality	Process, Resources

Table 4.22: Results from Topic 3 compared to the Periodic Table of BIM

Priority	Key Technical Terms from Topic Modelling	Technical Domains in the Periodic Table of BIM
1	Room	Process
2	Air	Process
3	System	Process
4	Temperature	Process
5	Fire	Process
6	Process	Process
7	Requirements	Process
8	Rooms	Process
9	Equipment	Process
10	Heat	Process

Table 4.23: Results from Topic 4 compared to the Periodic Table of BIM

Priority	Key Technical Terms from Topic Modelling	Technical Domains in the Periodic Table of BIM
1	Design	Foundation, Collaboration, Process
2	Construction	Foundation, Collaboration, Process
3	System	Technology, Enabling Tools
4	Process	Process
5	Programme	Resources, Digital Plan of Work stages
6	System	Process
7	Layer	
8	Quality	Process, Technology
9	Team	People, Performance
10	Time	Process

4.5.6 Lessons Learned from Experiment 2

Web scraping was performed to gather data related to Crossrail and BIM for the purpose of studying the use of BIM in enhancing performance in major construction project delivery. The collected data was analysed using TM techniques, including lemmatization, phrasing, stop word removal, and Document-Term Matrix creation.

The analysis resulted in a list of the most frequent words used in the corpus, as well as a visualization of topics created by the LDA model.

The results of the text analysis, shown in Table 4.13, indicate that the word “process” was the most frequent word, appearing 405 times, while performance-related words such as “efficiency”, “time”, “quality”, and “cost” were also mentioned frequently. The visualization of topics generated by the LDA model revealed that the topics were related to the different stages of construction, such as design, pre-construction, procurement, and commissioning, with the word “performance” being indicated as well. This suggests that BIM could play a role in performance enhancement during major construction project delivery.

In conclusion, the web scraping and text analysis performed provides a summary of the use of BIM in enhancing performance in major construction project delivery, particularly in the context of Crossrail. The results indicate that there are still challenges in improving efficiency, but the use of BIM has the potential to play a role in performance enhancement. In addition, the mixture of insights from BIMKG, the Periodic Table of BIM, and the BIMTAPE Framework brings a deeper understanding of the intricate relationships within the BIM landscape. This synergy reinforces the potential for well-informed decision-making and effective knowledge engineering in the context of major project delivery. The BIMTAPE Framework, with its emphasis on foundational aspects through TM, further amplifies the implications of the study for enhancing performance in significant construction projects. Together, these analytical tools offer a multifaceted perspective that combines textual analysis, the dynamics of knowledge graphs, and structured frameworks. This collective approach contributes to a comprehensive comprehension of the complex interplay within the realm of BIM.

4.6 Data Collection: BIM TS 3-6

4.6.1 Introduction

The aim of this section is to outline the data collection process for BIM TS 3 to BIM TS 6, with each dataset stored separately to enable independent TM analysis. Rather than presenting the data collection process for each dataset in separate sections, this combined approach allows the researcher to address data collection cohesively, avoiding unnecessary repetition. Each dataset (e.g., regulatory documents, standards, policy papers) is retained in its distinct format, allowing for independent TM analysis, comparison, and validation across different document types. This thesis encompasses three interrelated components, which include:

(1) An systematic literature review to understand the necessity of text analysis in addressing the challenges faced in BIM pervasive major project delivery. These challenges have been identified through lessons learned and pertain to important issues such as budget, quality, and schedule.
(2) A generic TM process framework to improve data collection and usage with the goal of enhancing performance.
(3) An examination of the findings from TM experiments, using the framework, to demonstrate its effectiveness in identifying and solving problems in BIM pervasive major project delivery.

Therefore, it is crucial to analyse these documents using TM techniques. In subsequent sections, the researcher presents the text analysis process for BIM TS 3, TS 4, TS 5, and TS 6, explaining the lessons learned, and summaries for each stage.

4.6.2 Body of Text: Data Collection TS3-TS6

4.6.2.1 BIM TS 3: Approved Documents

4.6.2.1.1 A Brief Description about Approved Documents

The approved documents outline what is typically considered acceptable for meeting the relevant requirements of the Building Regulations to which they are subject. Pertinent requirements and expectations are expounded by the Department for Levelling Up, Housing and Communities (DfLU, 2021), as explained in this subsection.

When the guidance provided in an approved document is followed, it is presumed that the requirements outlined in the guidance will be met. However, this does not guarantee compliance, as the guidance may not apply in situations where the case is unique or unusual in some way. It is important to note that there may be alternative methods to fulfil the requirements outlined by the Building Regulations, and there is no requirement to follow the solutions specified in an approved document. If a company wants to meet a relevant requirement in a different way than what is described in an approved document, it should consult with the relevant building control body.

Some approved documents include mandatory provisions that must be followed precisely, as specified by regulations or by the Secretary of State through methods of test or calculation. Each approved document pertains specifically to the requirements of the Building Regulations that the document addresses. However, building work must also adhere to any other applicable requirements of the Building Regulations. The legislative structure in the UK is divided into three jurisdictions, each with its own governing body and regulatory documents: England and Wales, Scotland, and Northern Ireland. Technical guidance documents are also available in Scotland and Northern Ireland. These documents not only provide clear prescriptive approaches but also offer performance-based solutions as a means of demonstrating compliance with the requirements of the Building Regulations (DfLU, 2021).

4.6.2.1.2 Data Collection for Regulation Documents

The TM analysis in this experiment was focused on evaluating the implementation of the requirements of approved documents in England and Wales. However, considering the similarities between the systems and building energy performance prediction (BEPP) tools utilized across the UK jurisdictions (Liu, 2007), it can be inferred that the findings from this experiment are likely to be relevant and applicable to all UK administrations.

To collect the approved documents, a Google search was conducted to retrieve the raw text PDFs of these documents. The majority of the documents were publicly accessible and there was no need to log in with university credentials to obtain them. The search terms used to access these documents were “building regulations and approved documents index" OR "Archived single Approved Documents". These documents are non-mandatory technical guidance documents that provide practical (quantified) solutions on how to comply with the regulations (DfLU, 2021). The research resulted in 26 Approved Documents (UK Government, 2021) which cover a wide range of topics such as structure, fire safety, energy efficiency, and fuel conservation standards (see Table 4.24). Although the number of documents may appear to be limited compared to other TM experiments, they occupy 78,714 KB of memory space in the system, indicating a substantial amount of textual data within these documents.

Table 4.24: Approved documents

Titles of Approved Documents	**Approved document code**	**No. of documents**
Structure	A	1
Fire safety	B	6
Site preparation and resistance to contaminates and moisture	C	1
Toxic substances	D	1
Resistance to sound	E	1
Ventilation	F	2
Sanitation, hot water safety and water efficiency	G	1
Drainage and waste disposal	H	1
Combustion appliances and fuel storage systems	J	1
Protection from falling, collision and impact	K	1
Conservation of fuel and power	L	2
Access to and use of buildings	M	2
Overheating	O	1
Electrical safety	P	1
Security in dwellings	Q	1
High speed electronic communications networks	R	2
Infrastructure for charging electric vehicles	S	1
Material and workmanship	T	1
Total		26

Source: UK Government (2021)

4.6.2.2 BIM TS 4 British Standards

The British Standard Institution (BSI, 2013) offers guidance on the adoption of BIM and its associated requirements for projects delivered using BIM (BSI, 2013). To gain a deeper understanding of the issues related to enhancing performance in major BIM project delivery, this research seeks to conduct a TM analysis of BSI documents from the UK, USA, and Europe. Accessing these documents requires a subscription, so the researcher utilized the University of Strathclyde's online access to the BSI, available from https://bsol-bsigroup-com.proxy.lib.strath.ac.uk/Home.

The researcher searched for documents containing the terms "BIM" and "building information modelling". 17 relevant documents were available for download, while others still required a subscription (Figure 4.26). Of the 17 available documents, 5 were drafts that should not be considered or used as official British Standards. These drafts, which are no longer current as of March 18, 2020, were issued to receive comments from interested parties and will be considered before publication (BSI, 2013). As a result, these drafts were excluded from the analysis. Table 4.25 provides a list of the names and titles of the documents that were included in the analysis.

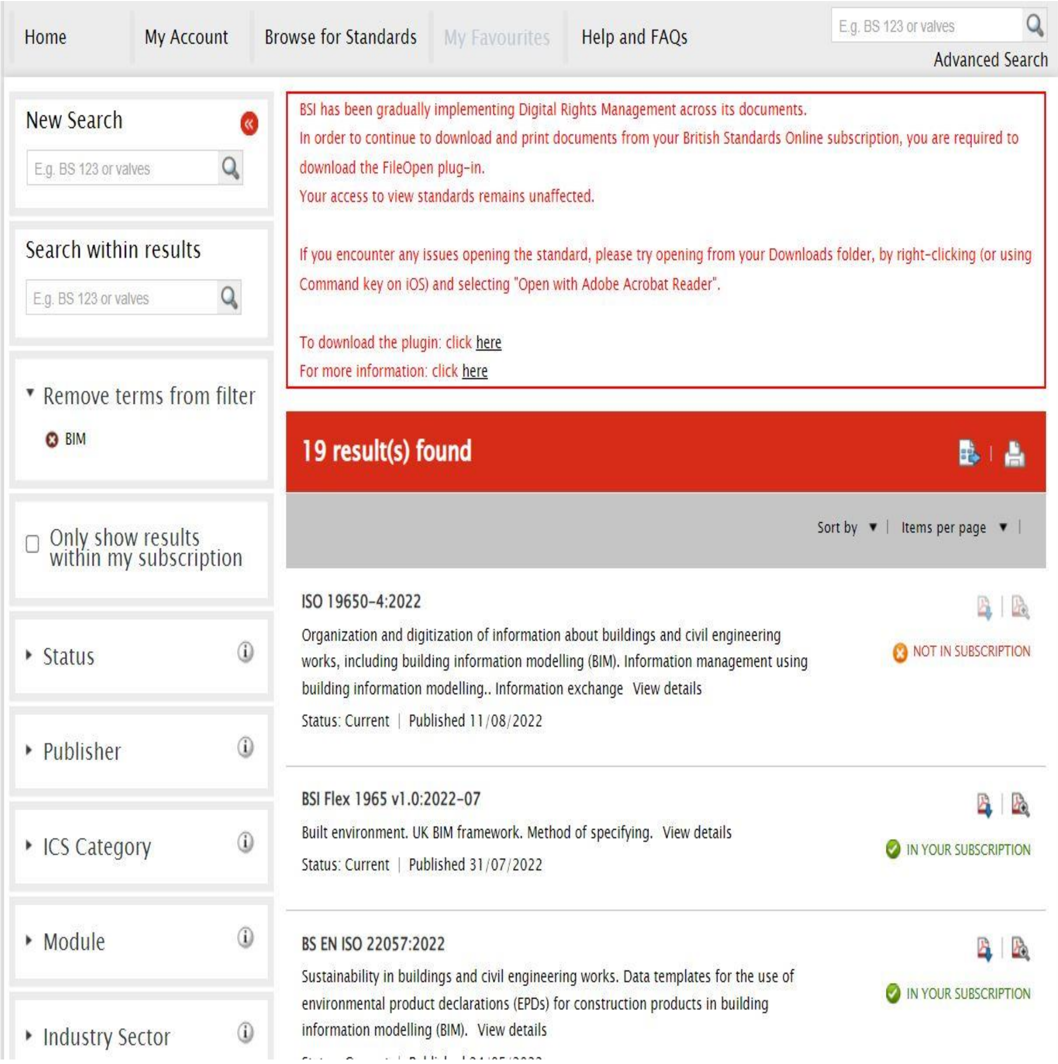


Figure 4.26: Example of some documents that are not in university subscription

Table 4.25: Standards names and the titles

No.	Codes of Standards	Titles of Standards
1	BS EN ISO 22057:2022	Sustainability in buildings and civil engineering works
2	PD ISO/TS 19166:2021	BIM to GIS conceptual mapping (B2GM)
3	PD ISO/TR 23262:2021	GIS (geospatial) / BIM interoperability
4	PD CEN/TS 17623:2021	BIM Properties for lighting — Luminaires and sensing devices
5	BS EN ISO 19650-3:2020	Organization and digitization of information about buildings and civil engineering works, including BIM
6	BS EN ISO 23387:2020	BIM. Data templates for construction objects used in the life cycle of built assets
7	BS EN ISO 19650-5:2020	Security-minded approach to information management
8	BS EN ISO 19650-1:2018	Information management using building information modelling — Part 1: Concepts and principles
9	BS EN ISO 19650-2:2018	Information management using building information modelling — Part 2: Delivery phase of the assets
10	PAS 1192-6:2018	Specification for collaborative sharing and use of structured Health and Safety information using BIM
11	PD ISO/TS 12911:2012	Framework for BIM guidance
12	BS EN ISO 19650-4:2022	Information management using building information modelling. Information exchange

4.6.2.3 BIM TS 5 National Policy Papers

National Policy Papers aim to enhance overall productivity by improving the quality and value for money of all construction projects through better government procurement practices, as declared by GCS 2011-2015. The strategy seeks to advance best practices within government and promote a unified voice to the construction industry. Additionally, it is focused on improving skills within government and the construction sector as a whole, while enhancing government's capacity and capability as a client. The strategy also strives to help the sector recruit and retain skilled workers.

Examining and analysing these strategies is essential for this study, as the research aims to investigate ways to enhance BIM performance in major construction project delivery. To this end, the researcher obtained the UK strategy documents by using the search term “UK Government Construction Strategy”. The search returned two documents: the Government Construction Strategy 2011-2015 (GCS 2011-2015) and the Government Construction Strategy 2016-2020 (GCS 2016-2020) (HM Government, 2011; HM Government, 2016). The researcher then performed TM on these documents, along with other relevant documents from previous sections.

The UK government and industry acknowledge that the country is not getting full value from public sector construction and that it has not fully utilized the potential of public procurement to drive growth, as acknowledged in GCS 2011-2015 (HM Government, 2011). This recognition led to a profound change in the relationship between public authorities and the construction industry, with the aim of ensuring that the government consistently receives a good deal, and the country gets the necessary social and economic infrastructure for the long term. GCS 2011-2015 aimed to make the public sector a better client by being more informed and better coordinated when specifying, designing, and procuring its requirements. This strategy also challenged industry business models and practices, replacing adversarial cultures with collaborative ones and demanding cost reduction and innovation within the supply chain to maintain market position and establish a bargaining position for the future.

GCS 2016-2020 sets out the plan for the government to improve its capability as a client in the construction industry and serve as a model for other clients. Building upon GCS 2011-2015, the 2016-2020 strategy aims to address the challenges posed by inflationary pressure in a rising market by driving increased construction productivity, which was expected to result in efficiency savings of GBP 1.7 billion over the course of the Parliament. This strategy also aims to evaluate and enhance the functional capability of central government as a client and improve understanding of the stages in programs where efficiencies can be achieved. Additionally, it seeks to consolidate the practices established in the previous strategy.

4.6.2.4 BIM TS 6 Enterprise Policy Papers

It is undisputed that the implementation of BIM provides value to the AEC industry and its clients. However, it still needs to be determined through field studies how BIM should best be introduced and supported. The effects of strategic policies implemented by a country, or a public client need to be better understood. There is evidence that a combination of mandatory BIM requirements from public and private clients and proactive adoption by the industry will accelerate the BIM adoption process (Lee and Borrmann, 2020). Further research is needed to determine the best way to achieve the optimal cost-benefit ratio and project goals from a BIM policy and management perspective.

With an increasing number of BIM projects being carried out globally, data-driven research can now be conducted to gain a deeper understanding of BIM policies and management factors (Lee and Borrmann, 2020). The systematic literature review presented in Chapter 2 revealed a need for a TM approach to enhance BIM performance in major construction project delivery. Hence, this section analyses BIM policy documents to gain valuable knowledge on how to enhance performance through the use of BIM.

4.6.2.4.1 BIM Policies

The government, being a major client in public construction projects, plays a crucial role in promoting BIM adoption within the construction industry, which is highly fragmented and heavily regulated. BIM policies can range from mandatory use of BIM in all publicly funded projects, with changes in legislation, if necessary, to providing financial and organizational support, to simple encouragement and support. Singapore, Finland, Korea, the USA, the UK, and Australia are at the forefront of BIM policymaking (Lee and Borrmann, 2020). In these countries, the government and its agencies have been instrumental in promoting and supporting the implementation of BIM but have adopted significantly different approaches in doing so.

The British government’s construction strategy, launched in 2011, aimed to reduce costs and lower the carbon footprint of construction projects through the use of BIM methods and technologies (Lee and Borrmann, 2020). The UK government also aimed to position the British construction industry as a leader in the digital construction era and make the UK a world leader in BIM in order to gain a competitive advantage in the international market (Maude, 2012). The key aspect of the 2011 UK construction strategy was to require "fully collaborative 3D-BIM", equivalent to BIM Level 2, for all centrally procured construction projects starting in 2016. As of the writing of this thesis, this goal has mostly been achieved, as evidenced by annual BIM surveys that show a significant increase in the adoption of BIM methods by the UK construction industry in recent years (Lee and Borrmann, 2020).

However, the current practice of using BIM execution plans has both potentials and limitations. One of these limitations is the limitations of human memory and analysis, which can be reduced by incorporating TM (Chen, 2022). Due to these limitations, there are problems with the current practice of using BIM execution plans. After reviewing key performance in major construction project delivery regarding delivery time, cost, and quality (as discussed in Chapter 2), it is clear that a comprehensive approach to evaluating BIM performance is needed. Therefore, BIM execution planning would be beneficial for major projects that heavily rely on BIM.

4.6.2.4.2 BIM Policies Collection

Several studies have underscored the importance of analysing BIM policy papers to improve BIM performance in large-scale projects. To achieve this objective, this study uses a search string to locate BIM policy papers at companies via Google. The search string incorporates the "intitle:" operator, which searches for web pages containing specific terms in their titles. The initial search generated a large number of results, with 4,580 relevant hits. However, to narrow down the search results to papers that were relevant to the study's focus, the researcher applied certain criteria, including specifying the document type required for text analysis and setting the desired format to PDF. By doing so, the researcher was able to select documents that concentrate on BIM policies. To locate PDFs related to BIM policies at companies, the search string used was "intitle:bim policy" OR "intitle:bim strategy" OR "intitle:bim implementation" filetype:pdf. By applying this string, the researcher identified 24 relevant PDF documents, which were subsequently downloaded for further analysis. The specifics of each downloaded BIM policy are summarized in Table 4.25.

Table 4.26: BIM policy papers

No.	Company names	Coverages of BIM policy
1	Anthony Keith Architects	3D visualisations to assist in the design process
2	CITYAXIS	Design, construction, and maintenance
3	Diane Butterworth	• Creating/placing BIM objects, adding/updating product data • Collaborating with contractors within a Common Data Environment (CDE)
4	ESS Modular	Providing design and build package
5	I and H Brown	Consultations
6	IFSE group	Design
7	IID Architects	Using Autodesk Revit platform for producing object-based design and construction models
8	ISG	Design and supply chain
9	James Engineering	Design
10	Lucas	3D software, information management, and production
11	Ogilvie Group	Consultations
12	ORMS	The use of Autodesk Revit on all projects from the very beginning of the design process
13	Pell Frischmann	Management, design, and coz\znstruction supervision
14	PIP	Production of 3D modelling
15	Queensland	The full lifecycle
16	Readie	Project and site management processes, supply chain
17	Rolton Group	Design
18	Scott Tallon Walker Architects	The use of Autodesk Revit on all projects from the very beginning of the design process
19	Sir Robert McAlpine	All project activities
20	Stephen George	Consultations
21	STEPNELL	Produce design works in 3D format
22	Streif	Consultations
23	Vinci Construction	Consultations
24	Wernick Buildings Ltd.	The full lifecycle

4.6.3 Managing Downloaded Data from TS 3 – TS 6

Having obtained BIM TS 3-6 documents in PDF format, TM modelling requires extracting plain text from these documents. The conversion of PDF documents to plain text format was necessary due to the memory space limitations in the system. PDF files tend to take up more space as they contain images, fonts, and references, etc. Careful consideration was given to the conversion of PDF documents to plain text, so as not to leave out any crucial information, which could potentially skew the results of the modelling. The researcher meticulously compared the plain text of each document with its original PDF counterpart to ensure accuracy. A PDF-to-text converter application was utilized to transform the PDFs into text files. The majority of the documents were successfully converted to plain text.

However, some documents faced challenges during the conversion process and could not be transformed into the desired plain text format. Several documents were secured with a password and could not be converted to plain text using the standard conversion process, as depicted in Figure 4.27. To overcome this issue, the researcher manually copied the text from these documents and pasted it into a word document, which was then converted to plain text, as shown in Figure 4.28. Additionally, some publishers saved the paper in an image format, causing difficulty in conversion from PDF to plain text.

The researcher encountered issues such as missing words or strange syntax when attempting to convert these papers. To address this, the researcher utilized an Optical Character Recognition (OCR) reader (Amazon Web Services, 2024) to convert the PDFs of these publications to plain text. Despite these efforts, a few documents still posed challenges and the researcher was unable to produce an accurate plain text version. The researcher then utilized the built-in Windows snipping tool to capture snapshots of the paper's text, saving them in image format, and using the OCR reader once more to generate the plain text files. Finally, the conversion process from PDF to plain text for all documents was completed.

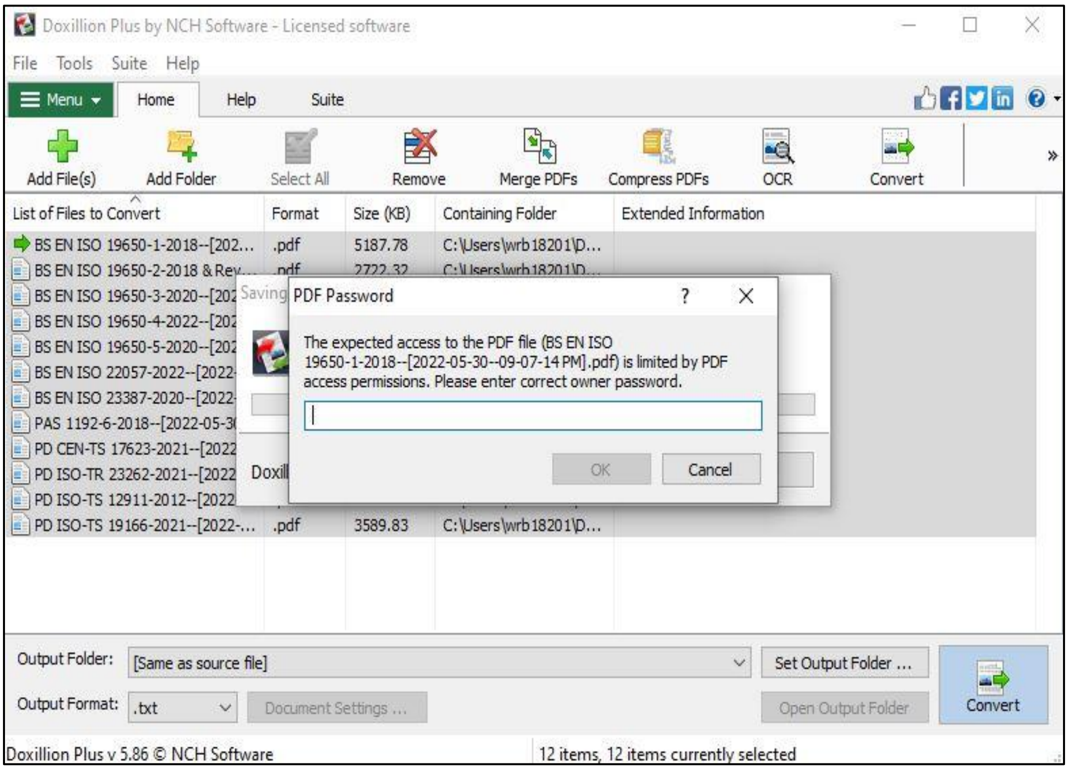


Figure 4.27: Encryption and protection of the document using password

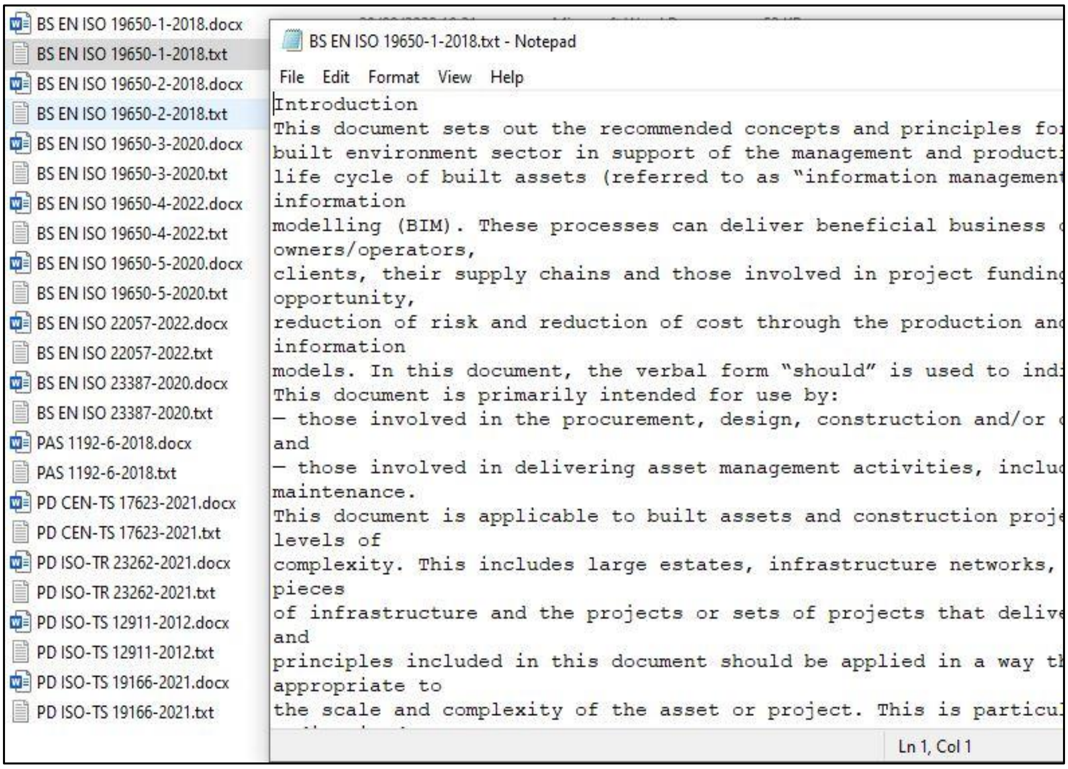


Figure 4.28: Converting word documents to plain text

The researcher then consulted with domain experts and decided to exclude the acknowledgement and reference chapters from the TM modelling process as they did not contain significant information. The researcher manually removed the acknowledgement and reference sections from the plain text of all 64 documents.

4.7 Experiment 3: TS 3 Approved Documents TM Process

4.7.1 Pre-Processing of Converted Text Data

In the previous section, all PDF documents were converted to plain text format to prepare them for analysis. The next step involves pre-processing this textual data to ensure it is ready for TM techniques. This pre-processing stage includes several key steps: removing unnecessary elements such as punctuation, numbers, and stop words; normalizing the text by converting it to lowercase; and applying stemming or lemmatization to reduce words to their base forms. These steps create a clean, consistent dataset, essential for accurate and efficient TM analysis. Pre-processing is crucial in TM as it helps to extract meaningful knowledge from unstructured text data (Nayak et al., 2016). However, Munková (2013) observed that pre-processing is the most time-consuming phase in TM when discovering sequential patterns from textual data. Having obtained the plain texts of the Approved Documents, TM modelling required extracting valuable information from these plain texts. However, the plain texts often contain irrelevant information such as punctuation marks, URLs, symbols, and numbers that can negatively impact the modelling results and slow down the processing time. To address this, NLP libraries were used to clean the text data. The data analysis and processing was conducted using RStudio programming language, and various NLP and other libraries like ggplot2, topic model, stemming, etc., were utilized to pre-process the textual data.

The researcher initiated the pre-processing of the plain text files. However, manually cleaning each of the 26 documents would have been unfeasibly time-consuming. To avoid this, the researcher utilized a TM library to automatically read all the files from a specific folder by setting the working directory (as shown in Figure 4.29). The researcher followed a well-established process, as used in previous experiments, to remove any unneeded data from the documents in order to maintain consistency. A brief overview of the steps taken during text cleaning is outlined below.

```
> cname <- file.path("C:\\Users\\wrb18201\\OneDrive - University of Strathclyde\\Approved documents")
> cname
[1] "C:\\Users\\wrb18201\\OneDrive - University of Strathclyde\\Approved documents"
> dir(cname)   # Use this to check to see that your texts have loaded.
 [1] "AD_Regulation_7.txt"
 [2] "AD_S.txt"
 [3] "ADB_Vol1_Dwellings_2019_edition_inc_2020_amendments_nosecurity.txt"
 [4] "ADB_Vol2_Buildings_other_than_dwellings_2019_edition_inc_2020_amendments_nosecurity.txt"
 [5] "ADD_LOCKED_nosecurity.txt"
 [6] "ADE_LOCKED1_nosecurity.txt"
 [7] "ADF1.txt"
 [8] "ADF2_revised.txt"
 [9] "ADL1.txt"
[10] "ADL2.txt"
[11] "ADO.txt"
[12] "Approved_Document_B__fire_safety__volume_1_-_2019_edition_nosecurity.txt"
[13] "Approved_Document_B__fire_safety__volume_1_dwellinghouses_nosecurity.txt"
[14] "Approved_Document_B__fire_safety__volume_2_-_2019_edition_nosecurity.txt"
[15] "Approved_Document_B__fire_safety__volume_2_buildings_other_than_dwellinghouses_nosecurity.txt"
[16] "Approved_Document_K_nosecurity.txt"
[17] "Approved_Document_M_vol_2_nosecurity.txt"
[18] "BR__PDF_AD__R__2016_nosecurity.txt"
```

Figure 4.29: Importing data to Rstudio for analysis

4.7.2 Transforming Text

The conversion of documents to a data frame and then to a corpus is the first step in the text transformation process. This corpus is a collection of text documents and can be processed using functions provided in the tm package (Feinerer, 2012). The corpus requires a few transformations, including converting letters to lowercase and removing punctuations, numbers, and stop words. In this case, the general English stop-words list was customized by adding “approved”, “documents”, “regulation”, and “documentations”. Additionally, hyperlinks were removed in the example (see Figure 4.30).

```
## Preprocessing
docs <- tm_map(docs,removePunctuation)
docs <- tm_map(docs, removeNumbers)
docs <- tm_map(docs, tolower)
mystopwords <- c(stopwords('english'), 'document')
mystopwords <- setdiff(mystopwords, c('regulations', 'documentations', 'approved', 'documents'))
docs <- tm_map(docs, removeWords, mystopwords)
docs <- tm_map(docs, removeWords, c("building", "information"))
docs <- tm_map(docs, stripWhitespace)
docs <- tm_map(docs, PlainTextDocument)
```

Figure 4.30: Function for removing punctuations, numbers, and stop-words

In the code presented, the tm_map() function serves as an interface for applying various transformations, referred to as ‘mappings’, to corpora. The get-transformations() function provides a list of transformations that can be utilized, with the most frequently used being removePunctuation(), PlainTextDocument(), removeNumbers(), removeWords(), stemDocument(), and stripWhitespace(). A function called removeURL() is also defined for the purpose of removing hyperlinks. This function matches strings that begin with “http” followed by any combination of alphabetic characters and digits using the regular expression pattern “http[[:alnum:]]”. The text.replace() function in Python is then employed to remove strings matching this pattern.

Despite applying the transformation process, certain “noisy” characters still remained in the documents, and had to be eliminated to improve the analysis results. To address this, a Python code (as seen in Figure 4.31) was utilized to remove these unwanted characters in each document separately. This process of document cleaning was repeated in all subsequent experiments. Following the removal of these unnecessary characters in Python, the researcher initiated another transformation task within RStudio.

```
file=open("Approved_Document_B__fire_safety__volume_1_-_2019.txt","r", encoding='utf-8', errors='ignore')
text=file.read()
```

```
text = text.replace(r"(\n|/n|â€|¢|â°c)","")
```

```
text = text.replace(r"https?://(www\.)?", '')
```

```
text=text.replace('http://','')
text=text.replace('https://','')
text=text.replace('www.','')
```

```
import re

text = re.sub(r"(@\[www-A-Za-z0-9]+)|([^0-9A-Za-z \t])|(\w+:\/\/\S+)|^rt|http.+?", "", text)

print(text)
```

```
        ONLINE VERSIONThe Building Regulations 2010Fire safetyAPPROVED DOCUMENTVolume 1 DwellingsRequirement B1 Means of war
ning and escapeRequirement B2 Internal fi re spread liningsRequirement B3 Internal fi re spread structure Requirement B4 Exte
rnal fi re spreadRequirement B5 Access and facilities for the fi re service Regulations 63 72 and 382019 edition   for use in
England         ONLINE VERSION                    ONLINE VERSIONMain changes in the 2019 editionThis volume of this approved do
cument supports requirements B1 to B5 of Schedule 1 to the Building Regulations 2010 as well as regulations 63 72 and 38 It t
akes effect on 30 August 2019 for use in EnglandThe main changes areApproved Document B has been redrafted to clarify its lan
guage and content in line with the Department s style guide for approved documents This edition of the approved documentrepla
ces the 2006 edition including all amendments There are no changes from the previous edition to the technical guidance within
Approved Document BAs well as furthering the use of plain English the document has been significantly restructured  The desig
n of blocks of flats has moved from volume 2 to volume 1  Guidance on the design of sprinkler systems has been consolidated t
o a new Appendix E  European fire classifications are provided within the main body of the document   with transposition to a
national classification provided in Appendix B  The guidance on external stairs has been consolidated  Fire safety informatio
n under regulation 38 has been moved from an appendix   into a new section  The guidance on insulating core panels has moved
from an appendix into the Wall   and ceiling linings section  The guidance on fire dampers and ventilation systems has been c
onsolidated                  ONLINE VERSION                          ONLINE VERSIONThe approved documentsWhat is an approved
document    The Secretary of State has approved a series of documents that give practical guidance about how     to meet the
requirements of the Building Regulations 2010 for England These approved documents     give guidance on each of the technical
parts of the regulations and on regulation 7 see the back of     this document The approved documents provide guidance for co
mmon building situations    It is the responsibility of those carrying out building works to meet the requirements of the
```

Figure 4.31: Python code to clean the corpus

4.7.3 New Knowledge (Building A Term-Document Matrix)

The construction of a term-document matrix involves mapping the relationship between terms and documents. This matrix represents each term as a row and each document as a column, with the entries being the frequency of the term's occurrence in the document. According to Lopez-Lira (2023), this matrix can also be presented in a document-term matrix format, which involves reversing the row and column positions. In the present study, the researcher utilized the function 'Term-Document Matrix' to construct a term-document matrix from the processed corpus, as illustrated in Figure 4.32.

```
> tdm <- TermDocumentMatrix(docs)
> tdm
<<TermDocumentMatrix (terms: 7932, documents: 26)>>
Non-/sparse entries: 27406/178826
Sparsity           : 87%
Maximal term length: 52
Weighting          : term frequency (tf)
```

Figure 4.32: Term-document matrix formation

The term-document matrix consists of 7,932 terms and 26 documents, as shown by the results. The matrix is quite sparse, with 87% of its entries being zero. The researcher then identified the terms that are most frequently used. With the help of the matrix, various TM tasks can be performed, such as clustering, classification, and frequency analysis (as shown in Figure 4.33). When there are too many terms, the size of the term-document matrix can be reduced by selecting terms that appear in a minimum number of documents, or by using the TF-IDF (term frequency-inverse document frequency) method (Wu et al., 2008).

```
> findFreqTerms(dtm, lowfreq=20)   # Change "50" to whatever is most appropriate for yo
ur data.
  [1] "ability"          "able"             "absorbent"        "absorption"
  [5] "accept"           "acceptable"       "accepted"         "access"
  [9] "accessible"       "accommodate"      "accommodation"    "accordance"
 [13] "according"        "accordingly"      "account"          "accreditation"
 [17] "accredited"       "achieve"          "achieved"         "achieves"
 [21] "achieving"        "acoustic"         "across"           "act"
 [25] "action"           "actions"          "activity"         "actual"
 [29] "added"            "addition"         "additional"       "address"
 [33] "addresses"        "adequate"         "adequately"       "adjacent"
 [37] "adjoining"        "adjusting"        "adopt"            "adopted"
 [41] "adoption"         "advice"           "affect"           "affected"
 [45] "agency"           "aggregate"        "air"              "airborne"
 [49] "airtightness"     "alarm"            "alarms"           "allow"
 [53] "along"            "also"             "alteration"       "alterations"
 [57] "alternative"      "alternatively"    "although"         "aluminium"
 [61] "always"           "amd"              "amended"          "amendment"
 [65] "amendments"       "amount"           "analysis"         "ancillary"
 [69] "angle"            "annex"            "annexes"          "another"
 [73] "appendix"         "appliance"        "appliances"       "applicable"
 [77] "application"      "applications"     "applied"          "applies"
 [81] "apply"            "approach"         "appropriate"      "appropriately"
 [85] "approval"         "approvedapproved" "april"            "architectural"
 [89] "area"             "areas"            "around"           "arrangement"
 [93] "arrangements"     "asfp"             "aspects"          "assemblies"
 [97] "assembly"         "assess"           "assessed"         "assessing"
[101] "assessment"       "assist"           "associated"       "association"
[105] "assume"           "assumed"          "assumptions"      "attached"
[109] "attention"        "authorities"      "authority"        "automatic"
```

Figure 4.33: Frequency of the terms

4.7.3.1 Frequent Terms and Associations

The researcher has analysed the commonly used words and the relationship between them. There are a total of 26 documents in the corpus. The code, findFrequentTerms(), identifies the most frequently used terms with a frequency of at least ten. The terms are sorted alphabetically, not by frequency or popularity. To visualize the top frequent terms, the researcher creates a bar plot of these terms. The frequency of the terms is determined by the researcher using the ‘rowSums’ function from the term-document matrix. The researcher then selects the terms that appear in all 26 documents and displays them in a bar plot using the ‘ggplot2’ package. Figure 4.34 illustrates this bar plot. The code below specifies a bar plot with the ‘geom=“bar’ parameter and swaps the x- and y-axis with coord_f(). After applying some transformations to eliminate sparse terms, a bar plot of the terms that appear at least 20 times in the corpus is displayed in Figure 4.35.

```
> p <- ggplot(subset(df, freq>20), aes(x = reorder(term, freq), y = freq)) +
+   geom_bar(stat = "identity") +
+   theme(axis.text.x=element_text(angle=90, hjust=1))
> p
```

Figure 4.34: A code to create a barplot of words that appears at least 20 times

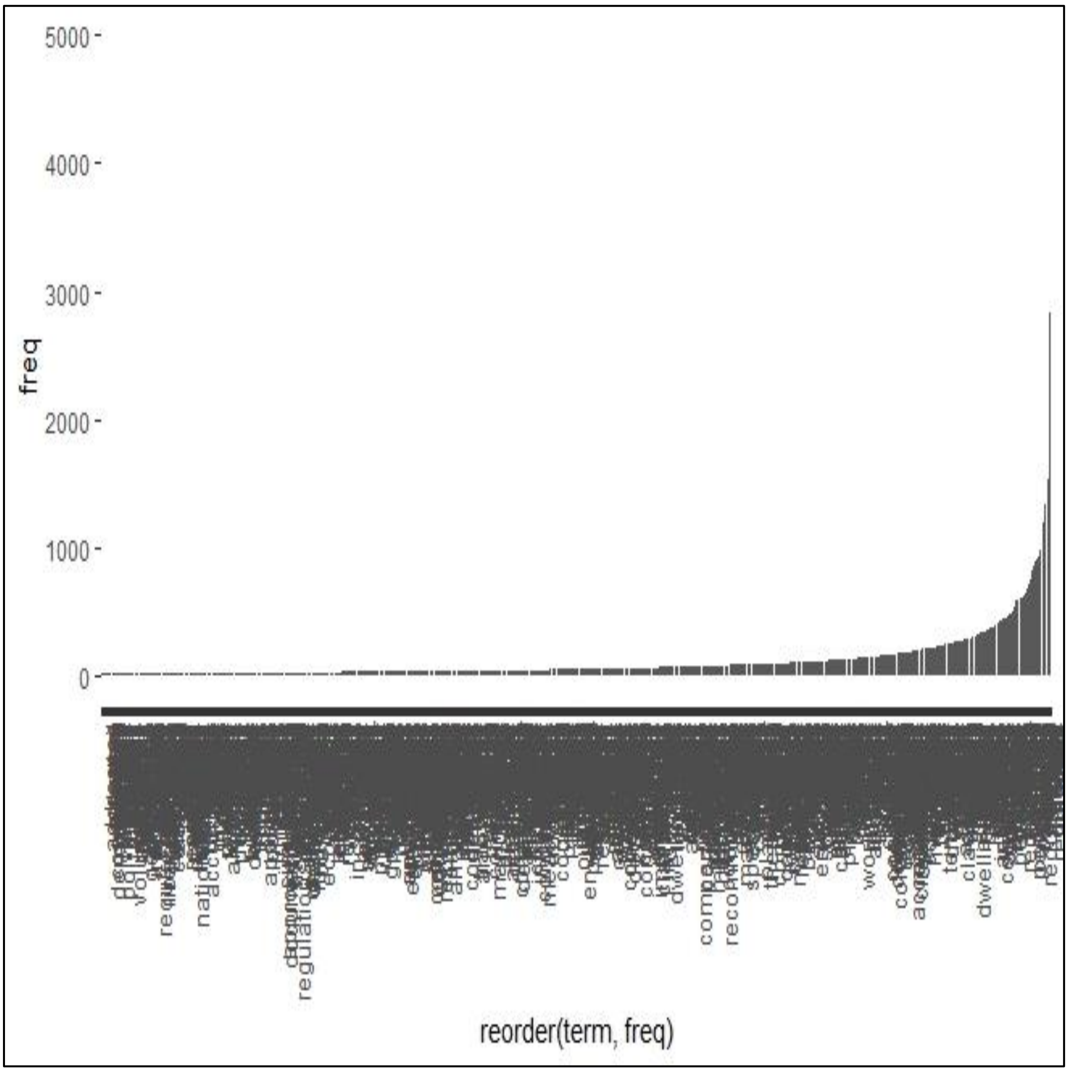


Figure 4.35: A barplot of words that appears at least 20 times

The bar plot of the frequently used words, that appear at least 20 times in the corpus, is challenging to interpret, as it includes a large number of terms from the term-document matrix, which comprises 7,932 terms. To address this issue, the researcher considered reducing the size of the term-document matrix by selecting terms that appear in a minimum number of documents or filtering terms with the TF-IDF (term frequency-inverse document frequency) method (Wu et al., 2008). As a result, the researcher chose to plot only 10% of the 7,932 terms (as seen in Figure 4.36), and Table 4.27 below displays the bar plot of terms that have a frequency of 20 or more.

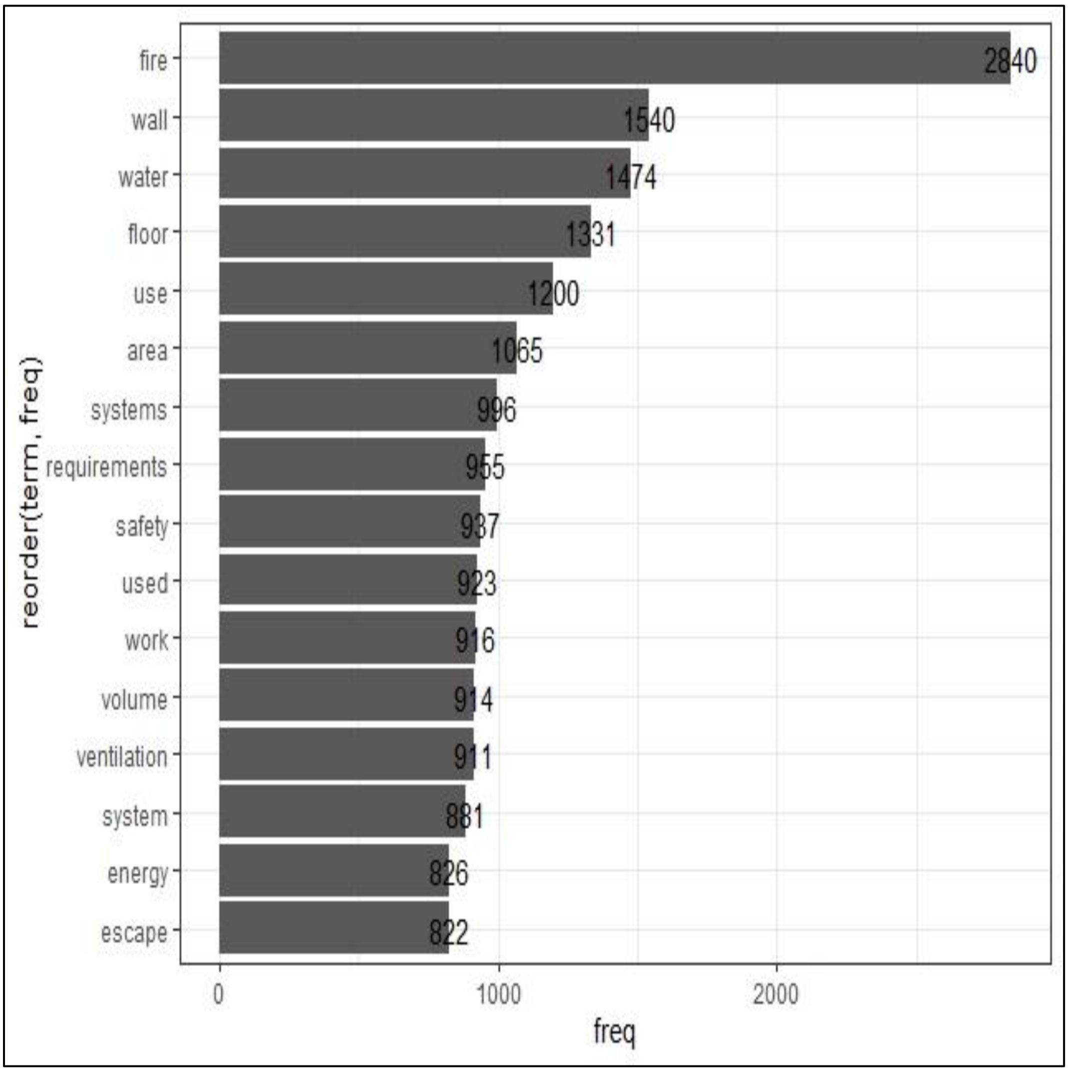


Figure 4.36: 10 % of 7,932 high-frequency words (20+ instances)

Table 4.27: Top 20 words

No	Words	Frequencies
1	Fire	2840
2	Wall	1540
3	Water	1474
4	Floor	1331
5	Use	1200
6	Area	1065
7	Systems	996
8	Requirements	955
9	Safety	937
10	Used	923
11	Work	916
12	Volume	914
13	Ventilation	911
14	System	881
15	Energy	826
16	Escape	822
17	Walls	782
18	Resistance	755
19	External	750
20	Design	737

From the bar plot and the accompanying table, it is evident that the word with the highest frequency is “fire” with a frequency of 2840 times. This is followed by “wall”, “water”, and “floor” with frequencies of 1540, 1474, and 1331, respectively. In comparison, words such as “resistance”, “external”, and “design” have lower frequencies of 755, 750, and 737 occurrences, respectively.

4.7.3.2 Association and Topic Modelling

The examination of relationships between phrases in text is of significance, as it enables the identification of words that frequently co-occur and those that do not. This can be accomplished through the use of techniques such as text data mining (TDM) or document-term matrix (DTM) by first evaluating the correlation between a particular term and another phrase. When the two terms appear together, the correlation score is assigned a value of 1, and when they do not, it is assigned a value of 0. The correlation score serves as a measure of the strength of the relationship between the terms in the sample. Although powerful pairwise phrase correlations exist, correlation values as low as 0.1 may still be observed due to term variability.

Tables 4.28, 4.29, 4.30, and 4.31 present the 50 correlation values of various words related to the concepts of “cost”, “time”, “quality”, “BIM”, and “performance”. The selection of these specific words was driven by the objective of the research, which aims to utilize TM for enhancing performance in BIM-based major construction project delivery. These terms hold a crucial significance in terms of improving performance, and thus, the researcher sought to identify terms that are highly associated with these key performance aspects as mentioned in the systematic literature review (i.e., time, cost, and quality) with a correlation coefficient of not less than 0.85. This was done by utilizing the findAssocs() function, as demonstrated in Figure 4.37, to determine these associations.

```
> s <- findAssocs(dtm, "cost", corlimit=0.85) # specifying a correlatio
n limit of 0.95
> write.csv(s, file="AssociationWords.csv")
```

Figure 4.37: Code for establishing association

Table 4.28: Terms associated with the word 'cost'

No.	Word	Correlation 0.85		No.	Word	Correlation 0.85
1	Electric	0.99		26	Earthed	0.96
2	Renovation	0.99		27	Elevation side	0.96
3	Charging	0.98		28	Envisaged	0.96
4	Infrastructure	0.98		29	Rent	0.96
5	Install	0.98		30	Expect	0.96
6	Vat	0.98		31	Feeder	0.96
7	Major	0.97		32	Fees	0.96
8	Acquisition	0.96		33	Financing	0.96
9	Asked	0.96		34	Footway	0.96
10	Avoids	0.96		35	Functionality	0.96
11	Backfilled	0.96		36	Future	0.96
12	Battlefields	0.96		37	Garden	0.96
13	Bays	0.96		38	Earthed	0.96
14	Boundary space	0.96		39	Elevation side	0.96
15	Buildings	0.96		40	Envisaged	0.96
16	requirement	0.96		41	Incoming	0.96
17	Busbar	0.96		42	Informing	0.96
18	Cable	0.96		43	Locating	0.96
19	Charge	0.96		44	Owned	0.96
20	Electric	0.99		45	Parking	0.96
21	Corresponds	0.96		46	Spaces	0.89
22	Deploying	0.96		47	Associated	0.87
23	Distribute	0.96		48	Remaining	0.87
24	Drawstrings	0.96		49	Coding	0.85
25	Dropped	0.96		50	Ownership	0.85

Table 4.29: Terms associated with the word 'time'

No.	Words	Correlation 0.85		No.	Words	Correlation 0.85
1	Using	0.98		26	Complies	0.89
2	Control	0.96		27	Determined	0.89
3	Achieving	0.95		28	Effective	0.89
4	Applies	0.95		29	High	0.89
5	Calculating	0.93		30	Individual	0.89
6	Fixed	0.93		31	Procedures	0.89
7	Higher	0.93		32	Secretary	0.89
8	Shall	0.93		33	Standards	0.89
9	Circumstances	0.92		34	Controlled	0.88
10	Ensure	0.92		35	Information	0.88
11	Reasonable	0.92		36	Measure	0.88
12	Building	0.91		37	Meet	0.88
13	Carrying	0.91		38	Referred	0.88
14	Operating	0.91		39	Results	0.88
15	Power	0.91		40	Section	0.88
16	Providing	0.91		41	Standardised	0.88
17	Stage	0.91		42	Suitably	0.88
18	State	0.91		43	Taken	0.88
19	Based	0.9		44	Central	0.87
20	System	0.9		45	Comply	0.87
21	View	0.9		46	Defined	0.87
22	Achieve	0.89		47	Element	0.87
23	Actual	0.89		48	Following	0.87
24	Cases	0.89		49	Improved	0.87
25	Using	0.98		50	Meeting	0.87

Table 4.30: Terms associated with the word 'quality'

No.	Words	Correlation 0.85
1	Comes	0.89
2	Counts	0.89
3	Mechanically	0.87
4	Settings	0.87
5	Activity	0.86
6	Comes	0.89

Table 4.31: Terms associated with the word 'performance'

No.	Word	Correlation 0.85		No.	Word	Correlation 0.85
1	Boarding	0.96		26	Parts	0.92
2	Data	0.96		27	Rating	0.92
3	Incorporating	0.95		28	Reduce	0.92
4	Sets	0.95		29	Specified	0.92
5	Flue	0.94		30	Ambient	0.91
6	Rather	0.94		31	Anticipated	0.91
7	Appropriate	0.93		32	Applicable	0.91
8	Cases	0.93		33	Assemblies	0.91
9	Considered	0.93		34	Described	0.91
10	Ensure	0.93		35	Ducts	0.91
11	Lower	0.93		36	Edge	0.91
12	Meeting	0.93		37	Made	0.91
13	Respectively	0.93		38	Recording	0.91
14	Used	0.93		39	Related	0.91
15	Around	0.92		40	Relates	0.91
16	Case	0.92		41	Sections	0.91
17	Either	0.92		42	Set	0.91
18	Elements	0.92		43	Standard	0.91
19	Included	0.92		44	Terms	0.91
20	Boarding	0.96		45	Test	0.91
21	Data	0.96		46	Additional	0.9
22	Incorporating	0.95		47	Amount	0.9
23	Sets	0.95		48	Guidance	0.9
24	Flue	0.94		49	Includes	0.9
25	Rather	0.94		50	Linings	0.9

The results from the analysis of the tables above demonstrate the associations between different terms in the documents. It is evident from the tables that words such as "electric", "renovation", "charging", and "infrastructure" are strongly linked to the term "cost", indicating a close relationship between these concepts. This could indicate that the implementation of electrical renovations, charging infrastructure, and the like, come at a cost to the project. The high association of words such as "using", "control", "achieving", "applies", and "calculating" with the term "time" highlights the importance of efficient and effective utilization of time in the project.

However, the results of the analysis show that there is a limited number of terms associated with the word "quality", with just 6 words in total. This may indicate that quality is not a significant consideration in the approved documents or that it is not specified in detail. Additionally, the findings suggest that the term "BIM" is not associated with any other words in the documents. This could imply that BIM has not been used in the approved documents, or it has not been mentioned in detail.

Furthermore, the researcher also found that the term "performance" is not associated with the term "BIM". This could suggest that the use of BIM has not been considered as a means of enhancing performance in the project delivery process. It could also indicate that there is a lack of recognition of the potential benefits of BIM in improving performance in major construction project delivery.

The analysis of the tables provides insights into the relationships between different terms in the approved documents. The results suggest that cost, time, and quality are significant considerations in the project, while the use of BIM and its impact on performance may not have been fully explored. Further research is needed to determine the reasons behind these findings and to explore the potential benefits of BIM in enhancing project performance.

4.7.3.3 Topic Modelling

The application of TM to the set of words in the text provides a deeper understanding of the various topics discussed. By using the LDA model, the researcher is able to identify the main themes and topics discussed in the text. The result of this analysis is presented in the form of top 10 topics and the top 5 words for each topic. Table 4.32 provides a visual representation of these topics and the related words. This information is valuable in understanding the key themes and areas of focus in the text, as well as how these topics are related to each other. Additionally, this information can be used to draw insights into the text, such as which topics are the most important, which topics are frequently discussed, and how topics are related to each other.

Table 4.32: Clusters of keywords in Topic modelling

Topic 1: Fire resistance and safety					
Keywords Cluster 1:	Fire	Wall	Materials	Safety	Resistance
Topic 2: Parking spaces					
Keywords Cluster 2:	Parking	Vehicle	Electric	Charge	Spaces
Topic 3: Air quality and ventilation					
Keywords Cluster 3:	Ventilation	Air	Volume	Area	Room
Topic 4: Indoor acoustics					
Keywords Cluster 4:	Floor	Per	Sound	Internal	Rooms
Topic 5: Thermal mass					
Keywords Cluster 5:	Wall	Design	Heat	Ground	Structures
Topic 6: Sound insulation					
Keywords Cluster 6:	Wall	Floor	Separating	Cavity	Sound
Topic 7					
Keywords Cluster 7:	Drainage	Water	Pipes	Sewer	Waste
Topic 8					
Keywords Cluster 8:	Energy	Heat	Dwelling	Systems	System
Topic 9					
Keywords Cluster 9:	Fire	Escape	Story	Safety	Floor
Topic 10					
Keywords Cluster 10:	Water	Hot	Safety	Work	Requirements

Table 4.32 shows the top 10 topics widely addressed ~~discussed~~ in the Approved Documents to further inform BIM pervasive design processes and design justification. Details about these topics are given below.

Topic 1 appears to be related to fire safety in buildings, specifically the materials used in constructing fire-resistant walls and the overall resistance of a building to fire. The presence of Keywords Cluster 1, which consists of “fire”, “wall”, “materials”, “safety”, and “resistance”, suggests that the documents are discussing ways to ensure the safety of buildings from fire hazards. BIM technology can simulate and analyse fire safety scenarios and help improve the fire resistance of building materials, making the building safer.

Topic 2 appears to be focused on electric vehicles and the provision of charging spaces for them. The Keywords Cluster 2, which consists of “parking”, “vehicle”, “electric”, “charge”, and “spaces”, indicate, that the documents are discussing the integration of electric vehicles into buildings, specifically the provision of parking spaces and charging infrastructure for them. BIM technology can be used to design and plan to park spaces, including those for electric vehicles, to improve parking efficiency and the charging infrastructure for electric vehicles.

Topic 3 appears to be related to the ventilation and air flow in buildings. The Keywords Cluster 3, which consists of “ventilation”, “air”, “volume”, “area”, and “room”, suggest, that the documents are discussing the design and implementation of air flow and ventilation systems in buildings, ensuring that the air quality is maintained and that the rooms are properly ventilated. BIM technology can simulate and analyse ventilation systems to improve air flow and air quality, making the building more comfortable and healthier for occupants.

Topic 4 appears to be related to soundproofing in buildings. The Keywords Cluster 4, which consists of “floor”, “per”, “sound”, “internal”, and “rooms”, indicate, that the documents are discussing the design and implementation of soundproofing measures in buildings, ensuring that sound does not travel between rooms or between different floors. BIM technology can simulate and analyse flooring systems to reduce the transmission of sound between rooms, improving the acoustic comfort of the building.

Topic 5 appears to be related to the design of building walls and their interaction with heat and the ground. The Keywords Cluster 5, which consists of “wall”, “design”, “heat”, “ground”, and “structures”, suggests that the documents are discussing the design of building walls and how they can effectively regulate heat transfer and maintain stability relative to the ground. BIM technology can simulate and analyse wall designs to improve heat transfer and energy efficiency, reducing the heating and cooling requirements of the building.

Topic 6 appears to be related to soundproofing between walls and floors in buildings. The Keywords Cluster 6, which consists of “wall”, “floor”, “separating”, “cavity”, and “sound”, indicates that the documents are discussing the design and implementation of soundproofing measures between walls and floors, ensuring that sound does not transfer between rooms or different floors. BIM technology can help optimize separating wall designs and cavity configurations to improve sound insulation and energy efficiency.

Topic 7 appears to be related to the drainage and waste management systems in buildings. The Keywords Cluster 7, which consists of “drainage”, “water”, “pipes”, “sewer”, and “waste”, suggests that the documents are discussing the design and implementation of drainage and waste management systems in buildings, ensuring that water and waste are properly removed from the building and disposed of. BIM technology can assist in the design and planning of drainage and waste management systems, improving the overall efficiency of these systems.

Topic 8 appears to be related to energy and heat systems in buildings. The Keywords Cluster 8, which consists of “energy”, “heat”, “dwelling”, “systems”, and “system”, indicates that the documents are discussing the design and implementation of energy and heat systems in buildings, ensuring that they are energy-efficient and that they provide the necessary heating and cooling. BIM technology can help optimize heating and cooling systems, reducing energy consumption, and improving the energy efficiency of the building.

Topic 9 appears to be related to fire safety in buildings, specifically the provision of fire escapes. The Keywords Cluster 9, which consists of “fire”, “escape”, “story”, “safety”, and “floor”, suggests that the documents are discussing the design and implementation of fire escapes in buildings, ensuring that there is a safe and effective means of escape in the event. BIM technology can assist in the design and planning of fire escape routes and fire safety systems, improving the overall safety of the building in case of a fire.

Topic 10 appears to be related to water supply and usage, to which BIM technology can help to improve the performance of water systems in building projects. With BIM, designers and engineers can simulate water systems in a building project to ensure that hot water requirements, safety standards and work requirements are met. This can help to reduce the risk of any water-related issues and increase efficiency in the building project (Liu et al., 2019).

BIM technology has become an increasingly important tool for enhancing performance and efficiency in the construction industry. Researchers have analysed various topics discussed in the approved documents and found that by leveraging the capabilities of BIM technology, they can address various challenges and improve various aspects of project delivery. For example, BIM technology can be used to simulate and analyse building energy systems, which can result in increased energy efficiency, improved safety, and better compliance with regulations and requirements. Additionally, by utilizing BIM technology, the accuracy and efficiency of project planning, construction, and management can be improved, resulting in more efficient use of resources and reduced project delivery time. These benefits demonstrate the importance of considering the role of BIM technology in enhancing performance and efficiency in major construction projects.

4.7.3.4 Word Cloud

The researcher utilized the “wordcloud” package (Fellows, 2012) to display the significance of words in the data obtained from the term-document matrix and topic modelling. First, the researcher transformed the term-document matrix into a standard matrix and calculated the frequency of each word (as shown in Figure 4.38). Next, the researcher used the “wordcloud” function to create a visual representation of the data. The first two parameters in the function provided a list of words and their frequencies. Words with a frequency below three were excluded from the plot as specified by the “min.freq=50” parameter. The “random.order=F” parameter ensured that the most frequent words were plotted first, making them appear at the centre of the cloud. Finally, the researcher set the colours of the plot according to frequency levels.

Figure 4.38: Word cloud

The researcher utilized the word cloud as a visual representation to showcase the significant words derived from the term-document matrix and topic modelling. The results confirmed that the approved documents contain information related to BIM, particularly regarding fire, wall, water, and floor. Moreover, the analysis highlights the significance of energy efficiency and sustainability, as words such as “energy”, “performance”, “design”, “air”, “materials”, and “systems” were also frequently used. BIM provides benefits for subcontractors by allowing them to anticipate building behaviour and by storing critical information about the building. Additionally, the frequent usage of words related to “dwelling systems” and “work requirements” indicate a focus on safety while using hot water. The final use of the building is also hinted at through the usage of words “floor” and “walls”. To further organize the data, the researcher applied hierarchical clustering to identify clusters of related words and remove sparse terms to avoid a cluttered plot.

4.7.3.5 Clustering Words

To further analyse the data obtained from the word cloud, hierarchical clustering techniques were applied to the terms. The first step in this process is to calculate the distances between terms using the ‘dist()’ function, which takes into account the scaled term frequencies. The ‘hclust()’ function is then used to cluster the terms and form a dendrogram, which is a tree-like diagram that shows the relationships between the terms.

The dendrogram is cut into 10 clusters, which is a commonly used number in hierarchical clustering analysis, as it provides a good balance between the number of terms in each cluster and the overall structure of the dendrogram. The agglomeration method used in this analysis is ‘ward.D2’, which is a method that seeks to minimize the variance of the distances between the newly merged clusters. This is achieved by computing the sum of squares of differences between the centroids of the two clusters before merging.

There are several other options for agglomeration methods in hierarchical clustering, including ‘single linkage’, ‘complete linkage’, ‘average linkage’, ‘median’, and ‘centroid’. Each of these methods has its own strengths and weaknesses and is best suited to different types of data. Further details on these methods and the advantages and disadvantages of each can be found in data mining textbooks, including the following: “Data Mining: Concepts and Techniques” (Han and Kamber, 2000); “Data Mining: Practical Machine Learning Tools and Techniques” (Hand et al., 2001); and “Data Mining: Practical Machine Learning Tools and Techniques” (Witten and Frank, 2005).

The code in Figure 4.39 was utilized to identify the clustered words. The researcher then splits the clustered words into ten distinct groups and creates tables to display the results, which can be seen in Figures 4.40 and 4.41. The dendrogram in Figure 4.40 displays the topics present in the documents by grouping related words together. For instance, words such as “appliance”, “back”, “collision”, etc. are grouped in one cluster, while another cluster contains words such as “also”, “materials”, "performance", etc. To further understand the contents of the documents, the researcher applied k-means and k-medoids algorithms to cluster the documents.

```
> ### K-means clustering
> dtmss <- removeSparseTerms(dtm, 0.95) # This makes a matrix that is only 15% empty space, maximum.

> dtmss
<<DocumentTermMatrix (documents: 24, terms: 4500)>>
Non-/sparse entries: 23655/84345
Sparsity           : 78%
Maximal term length: 24
Weighting          : term frequency (tf)
> m2 <- as.matrix(dtmss)
> write.csv(m2, file="TermDocMatrix.rdata")
> d <- dist(t(m2), method="euclidian")
> fit <- hclust(d, method = "ward.D2")
> plot(fit)
```

Figure 4.39: Code for clustering words

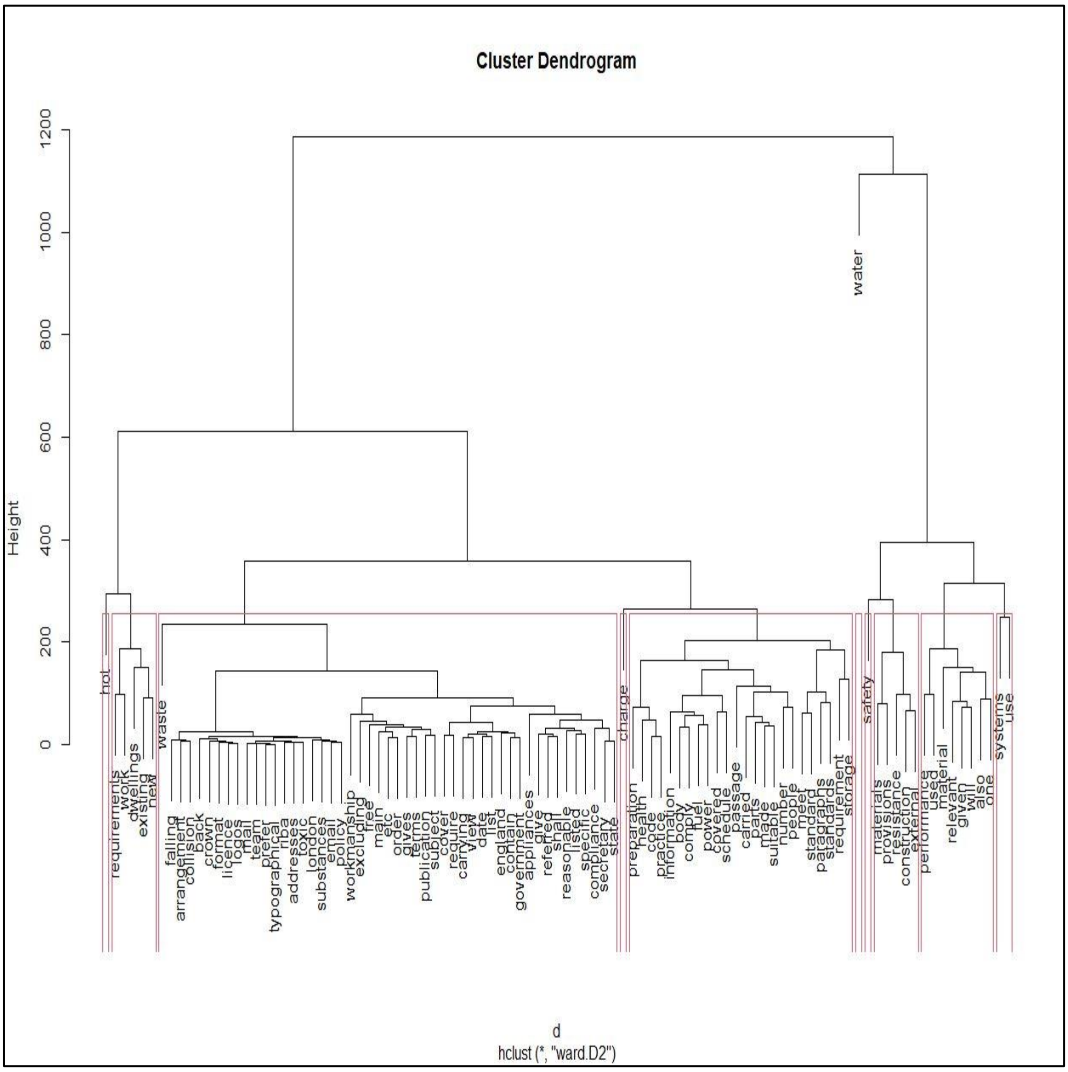


Figure 4.40: Dendrogram for group of words

```
> group <- (groups <- cutree(fit, k=10))
> group
  addresses         also   appliances  arrangement         back         body 
          1            2            1            1            1            3 
    carried     carrying       charge         code    collision   compliance 
          3            1            4            3            1            1 
     comply construction      contain        cover      covered        crown 
          3            5            1            1            3            1 
       date    dwellings        email      england          etc    excluding 
          1            6            1            1            1            1 
   existing     external      falling       format         free         fuel 
          6            5            1            1            1            3 
       give        given        gives   government       health          hot 
          1            2            1            1            3            7 
information      licence         list       listed        logos       london 
          3            1            1            1            1            1 
       made         mail         main     material    materials         meet 
          3            1            1            2            5            3 
        new       number          one        order   paragraphs        parts 
          6            3            2            1            3            3 
    passage       people  performance       policy        power     practice 
          3            3            2            1            3            3 
     prefer  preparation   provisions  publication   reasonable     referred 
          1            3            5            1            1            1 
   relevant      require  requirement requirements   resistance         riba 
          2            1            3            6            5            1 
     safety     schedule    secretary        shall     specific     standard 
          8            3            1            1            1            3 
  standards        state      storage      subject   substances     suitable 
          3            1            3            1            1            3 
    systems         team        terms        toxic typographical          use 
          9            1            1            1            1            9 
       used         view        waste        water         will         work 
          2            1            1           10            2            6 
workmanship 
          1 
```

Figure 4.41: Table of groups for clear view

4.7.3.5.1 Clustering Documents with the K-Means Algorithm

The researcher utilized the k-means clustering method in order to extract insights from the document set. The k-means algorithm takes the values in the matrix as numerical, as depicted in Figure 4.42. The term-document matrix was transformed into a document-term matrix for this purpose. The documents were then clustered using the ‘k-means()’ function with the number of clusters set to eight. The researcher analysed the most frequently occurring words in each cluster and also evaluated the cluster centres. It is worth mentioning that a fixed random seed was set using the ‘set.seed()’ function before executing the ‘k-means()’ function, in order to ensure reproducibility of the clustering results. This step is purely for the convenience of the thesis writing and is not required for readers to implement in their own code.

```
> m3 <- m2
> set.seed(122)
> k <- 8
> kmeansResult <- kmeans(m3, k)
> round(kmeansResult$centers, digits = 3)
  addresses    also appliances arrangement  back   body carried carrying
1     2.000  31.000     43.000       6.000 6.000  8.000  24.000    8.000
2     1.333  16.833      4.833       1.167 1.833 14.500   9.333    3.833
3     2.000 211.000     30.000      12.000 2.000 14.000  32.000    6.000
4     2.000  70.000     11.000       3.000 5.500 11.500  35.500    6.500
5     1.000  49.000     32.000       2.000 1.000  2.000  35.000    8.000
6     1.000   7.000      2.000       2.000 1.000 18.000   7.000    2.000
7     1.000  15.500     16.500       1.500 4.000 43.500  34.000   19.500
8     0.083   0.167      0.083       0.083 0.083  0.083   0.333    0.083
   charge   code collision compliance comply construction contain  cover
1   1.000 33.000     1.000     28.000 14.000        6.000   9.000  1.000
2   1.833 18.333     1.167      8.333 14.167       22.333   4.667  6.167
3   3.000 43.000     9.000     20.000 39.000      241.000  14.000  8.000
4   3.000 13.000     2.500     15.500 12.000      109.500   7.500  7.500
5   3.000 14.000     1.000      3.000  6.000       41.000   5.000 24.000
6 193.000  9.000     1.000      6.000 14.000        5.000   4.000  5.000
7   3.500  8.500     1.000     19.000 31.500       28.500   9.500  6.000
8   0.083  0.333     0.083      0.083  0.083        0.083   0.083  0.083
  covered crown   date dwellings email england    etc excluding existing
1   8.000 2.000  4.000    59.000 2.000  14.000 15.000     7.000    7.000
2   7.000 2.333  3.167    28.000 3.667   9.333  5.167     2.833   16.500
3  38.000 4.000  1.000    29.000 2.000   3.000 28.000    28.000   21.000
4  12.000 2.500  4.000    15.500 2.000   5.000 14.500    28.000   31.500
5  10.000 4.000  4.000    28.000 3.000   9.000 14.000     4.000   17.000
6  65.000 2.000  2.000    83.000 2.000   7.000  8.000     5.000    3.000
7  15.000 2.000 13.000   103.500 2.000  16.500  7.500     6.000  132.500
8   0.167 0.083  0.083     0.083 0.167   0.333  0.167     0.167    0.167
  external falling format   free   fuel  give   given  gives government
1    7.000   3.000  1.000  4.000 13.000  7.00  23.000  3.000      7.000
2   23.667   2.500  3.167  8.167  6.333  7.00  26.667  6.500      5.000
3  253.000  13.000  3.000 31.000 11.000 16.00 143.000 23.000      6.000
4  131.500   4.500  3.000  4.000 13.500 12.50  44.000 15.000      3.500
5    5.000   2.000  1.000 11.000  6.000 19.00  16.000  9.000      4.000
6    2.000   1.000  3.000  4.000  4.000  5.00   6.000  3.000      2.000
7   38.500   2.500  4.000  3.500 53.000 17.00  48.000  3.000     10.000
8    0.083   0.083  0.083  0.083  0.417  0.25   0.083  0.083      0.167
  health     hot information licence   list listed logos london   made
1 29.000 245.000      18.000   3.000  3.000  9.000 2.000  1.000 22.000
2 18.500   2.833      22.667   2.667  1.833  6.333 2.333  3.500  6.833
3 47.000   6.000      49.000   3.000  5.000 26.000 3.000  1.000 52.000
4 21.000   4.000      19.000   3.000  3.500 13.500 2.500  1.000 30.500
5 36.000   3.000      31.000   3.000  2.000  7.000 2.000  2.000 32.000
6  3.000   1.000      14.000   3.000  1.000 10.000 3.000  1.000  4.000
7  2.500  83.000      46.000   3.000 18.500 15.000 3.000  1.000 27.000
8  0.583   0.083       0.333   0.167  0.083  0.083 0.167  0.083  0.250
   mail   main material materials   meet    new number     one order
1 1.000  3.000   23.000     5.000  6.000 37.000 28.000  32.000  8.00
2 1.167  5.167   17.500    20.833 14.667 11.333  8.833  17.500  4.00
3 1.000 44.000  111.000   221.000 83.000 32.000 94.000 207.000 34.00
```

Figure 4.42: K-means algorithm

The researcher then utilized a code as depicted in Figure 4.43 to identify the top five words in each cluster. This information is then presented in Table 4.33, providing insights into the content of each cluster.

```
> for(i in 1:k){
+   cat(paste("cluster",i,": ", sep = ""))
+   s <- sort(kmeansResult$centers[i,], decreasing = T)
+   cat(names(s)[1:3],"\n")
+   #printtweetsinclusteri
+   #print(rdmDocuments[pamResult$clustering==i])+
+ }
```

Figure 4.43: A code to check the top five words in every cluster

Table 4.33: Top five keywords in clusters

Keywords Clusters	**Strings of Keywords**				
Cluster 1	Fire	Wall	Safety	Resistance	Escape
Cluster 2	Water	Hot	Safety	Supply	Sanitary
Cluster 3	Wall	Floor	Sound	Separating	Cavity
Cluster 4	Ventilation	Design	Wall	Water	Drainage
Cluster 5	Parking	Vehicle	Electric	Charge	Spaces
Cluster 6	Fire	Escape	Safety	Story	Floor
Cluster 7	Energy	Heat	Dwelling	Systems	System
Cluster 8	Area	Work	Requirements	Residential	Edition

The analysis of the top words and centres has revealed that the clusters are predominantly homogeneous in topic.

Cluster 1 deals with the fire protection measures that are used to ensure the safety of escape routes and provide a secure refuge. These measures are typically a combination of passive and active protection strategies, such as walls, floors, and doors, which provide robust passive fire protection.

Cluster 2 focuses on the issues related to sanitation, hot water safety, and water efficiency. Sanitation is a critical aspect of building design, as it helps maintain hygienic conditions and prevents the spread of disease. Similarly, hot water safety is crucial for ensuring that hot water systems in buildings are designed and installed in a manner that minimizes the risk of scalds and burns. Finally, water efficiency refers to the use of water in an efficient and sustainable manner, reducing waste and conserving resources.

Cluster 3 deals with sound-resisting construction, which is an important aspect of building design that helps reduce the transmission of noise between different parts of the building or between the building and its surroundings. This helps to create a quieter and more comfortable living environment.

Cluster 4 focuses on wall and floor ventilation, which is a requirement stipulated in Part C of the Building Regulations. This aspect of building design is crucial in preventing the build-up of stagnant air, mould, and potentially explosive gases. By ensuring proper ventilation, the indoor air quality of a building can be maintained, providing a healthier and safer living environment.

Cluster 5 pertains to the requirement for new homes and existing homes undergoing substantial renovations to have facilities for electric vehicle charging. This requirement is specified in UK Building Regulations Part S, and it is aimed at promoting the use of electric vehicles and reducing reliance on fossil fuels.

Cluster 6 is similar to Cluster 1, focusing on fire protection measures. Cluster 7 deals with efficient heating systems for low-energy homes, and Cluster 8 outlines the technical requirements applicable to building work that are necessary to protect the public interest.

The next step in the analysis process involves evaluating the results obtained from the k-medoids algorithm to gain a deeper understanding of the relationships between the various aspects of building construction and design.

4.7.3.5.2 Clustering Documents with the K-medoids Algorithm

In this section, the researcher employs the Partitioning Around Medoids (PAM) algorithm, a variant of k-medoids, for clustering purposes (Hennig, 2010). Unlike k-means, which utilizes means to represent clusters, k-medoids utilizes medoids, or representative objects, as a means of representation, making it more resilient to noise and outliers. The researcher utilizes the ‘pamk()’ function from the ‘fpc()’ package (Hennig, 2010) to implement the PAM algorithm and determine the number of clusters via the optimum average silhouette. To visualize the quality of clustering, a silhouette plot is generated. The code implementation and resulting outputs are depicted in Figure 4.44. The following code, shown in Figure 4.45, was utilized to plot the cluster, as demonstrated in Figure 4.46.

```
> library(fpc)
Warning message:
package 'fpc' was built under R version 4.0.5
> pamResult <- pamk(m3, metric='manhattan')
> k <- pamResult$nc
> pamResult <- pamResult$pamobject
> for (i in 1:k) {
+     cat("cluster", i, ": ",
+       colnames(pamResult$medoids)[which(pamResult$medoids[i,]==1)], "\n ")
+ }
cluster 1 :  achieve construction policy requirements systems within
 cluster 2 :  management policy team
 cluster 3 :  date design development policy work working
```

Figure 4.44: K-medoids algorithm

```
> layout(matrix(c(1, 2), 1, 2)) # set to two graphs per page
> plot(pamResult, color = F, labels = 4, lines = 0, cex = .8, col.clus = 1,
+      col.p = pamResult$clustering)
> layout(matrix(1))
```

Figure 4.45: Code for plotting the cluster of the words

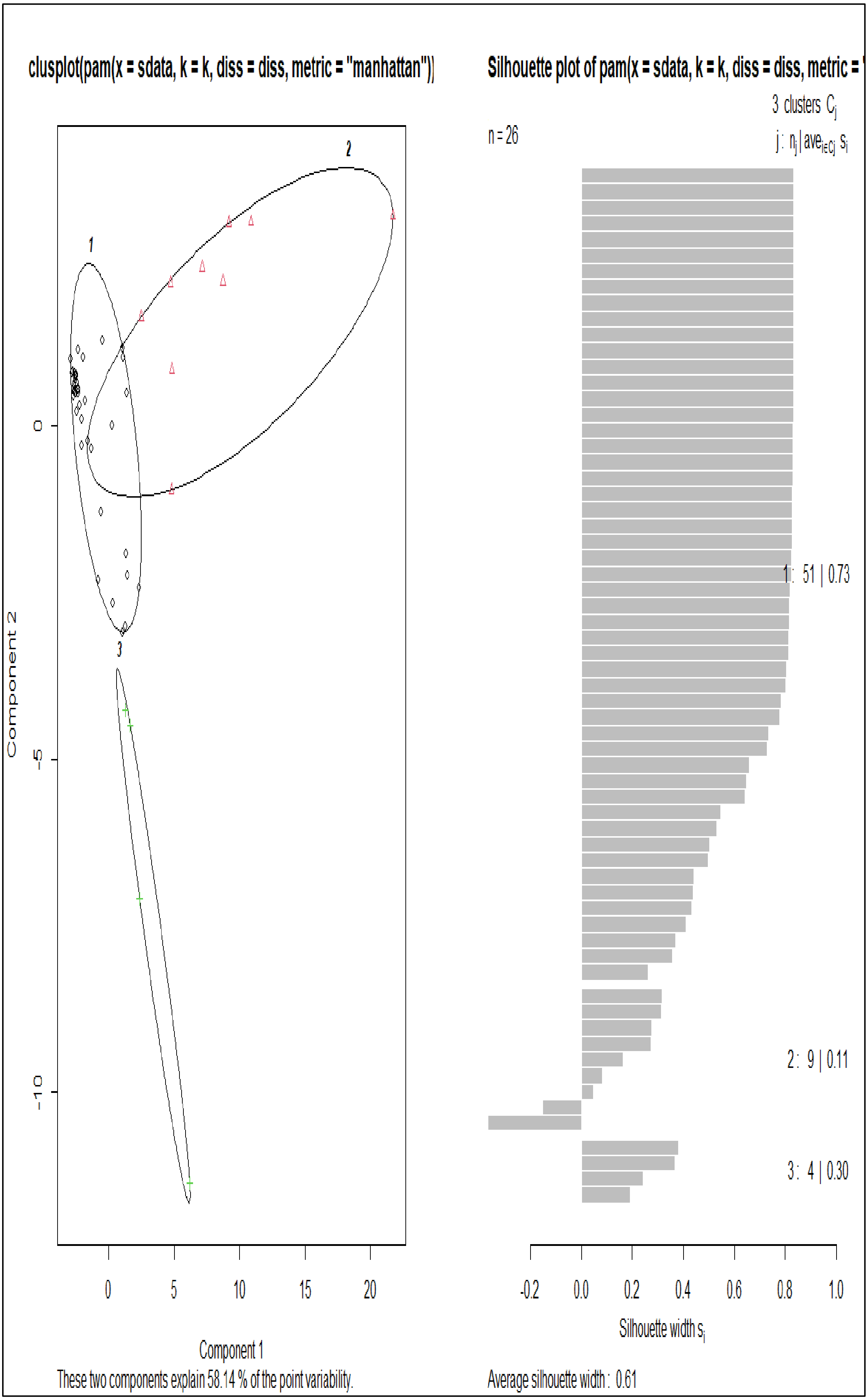


Figure 4.46: Clustering chart

The clustering analysis performed in this section reveals the presence of three distinct clusters of documents. Upon further examination of the results, it can be observed that clusters 2 and 3 are well-separated groups, each with a clear and distinct focus on a specific topic. The presence of a large overlap between clusters 1 and 2 can be understood by examining their respective medoids.

The use of silhouettes in this analysis provides additional insight into the quality of clustering. A high silhouette width (close to 1) indicates that the observations are well-grouped and properly assigned to their respective clusters. In contrast, a low silhouette width (near 0) suggests that the observation is situated between two clusters, potentially leading to ambiguity in its assignment. Negative silhouette width, on the other hand, indicates that the observation may have been misassigned to the incorrect cluster.

The average silhouette width of 0.61 suggests that the clusters are well separated from each other, although some observations within cluster 3 exhibit negative silhouette width. This suggests that these observations may be better suited for other clusters and may warrant further investigation.

Overall, the results and charts presented in this section provide valuable information about the cluster structure and quality of the clustering analysis performed. These insights can be utilized in further analysis and decision making.

4.7.4 Lessons Learned from Experiment 3

Lessons learned from the experiments conducted include the importance of analysing frequently used words and their relationships in a corpus of documents. The experiment revealed that visualizing the top frequent terms through bar plots can help understand the most commonly used words. However, when dealing with a large number of terms, reducing the size of the term-document matrix by selecting specific terms or using TF-IDF can improve interpretability. Furthermore, the analysis of associations between terms provided insights into the relationships between different concepts such as cost, time, quality, and BIM.

It was observed that certain concepts like cost and time had strong associations with specific terms, while others like quality and BIM had limited associations. These findings highlight the need for further research to explore the reasons behind these associations and the potential benefits of BIM in improving project performance. Additionally, applying topic modelling techniques, such as LDA, allowed for the identification of main themes and topics discussed in the text, providing a deeper understanding of the text's content and its interconnections. These findings emphasize the significance of analysing text data to gain insights and inform decision-making in various domains.

In addition, the experiment demonstrates the effectiveness of utilizing word clouds and clustering techniques to gain insights from textual data. By visualizing word frequencies and creating word clouds, the researcher was able to identify significant keywords and topics present in the data. The clustering analysis further enhanced the understanding of the relationships and groupings within the data by organizing similar terms and documents into clusters. The use of different clustering algorithms, such as k-means and k-medoids, provided different perspectives on the data structure and helped identify distinct clusters with specific themes. Additionally, the evaluation of silhouette widths offered a measure of the quality of clustering, indicating the degree of separation and appropriateness of cluster assignments. These techniques and analyses serve as valuable tools for exploring and understanding large textual datasets, providing researchers with meaningful insights, and facilitating decision-making processes.

4.8 Experiment 4: TS 4 British Standards TM Process

4.8.1 Pre-Processing of the Raw Textual Data

Approximately 12 documents were obtained from the BSI website to support the TM process for TS 4 British Standards, as outlined in Section 4.6.2.2. This relatively limited number of documents was due to the fact that the researcher was only able to successfully download 17 related to BIM, with five being excluded. Additionally, access to further documents was restricted as they required a subscription.

The focus of the experiment was to outline the methodology utilized by the researcher to conduct text analysis and topic assignment within RStudio. This process began with the cleaning of the text by eliminating embedded URLs and irrelevant terms. The corpus was then transformed by converting all letters to lowercase and removing punctuation, numbers, and commonly occurring words known as stop words. The standard English stop-word list was modified to include the terms “BSI”, “building”, “information”, “modelling”, and “BIM” as these words were likely to appear frequently in the corpus and could potentially affect the accuracy of the results. In the second step, the researcher tokenized the corpus to identify the smallest meaningful units of words and combined them into phrases. This minimizes the impact of repetitive language tokens, such as articles and prepositions, on the accuracy of the NLP results by ensuring that the input text accurately reflects the content.

The third step involved the creation of a dictionary of words from the parsed texts and the production of a corpus, which served as parameters for topic estimation. In the fourth step, the researcher constructed a term-document matrix from the processed corpus using the “Term-Document Matrix()” function (Shin and Bulut, 2021). The researcher then examined the most frequent words and the relationships between words by exploring the connections between phrases in the text.

Finally, the researcher calculated topic probabilities and categorized the topic words using the dictionary and corpus created earlier. This was achieved using the LDA method developed by Blei et al. (2003).

4.8.2 New Knowledge

4.8.2.1 Building a Term-Document Matrix

As depicted in Figure 4.47, the term-document matrix comprises 7,058 terms and 12 documents. The matrix exhibits a high degree of sparsity, with 84% of its entries being zero. The researcher then conducted an analysis of the most frequently used terms, which can be viewed in Figure 4.48. This matrix can serve as the basis for a variety of data mining tasks, including clustering, classification, and association analysis.

When the number of terms in the matrix becomes excessive, it can be challenging to perform data mining tasks effectively. To address this issue, feature selection techniques can be employed to reduce the size of the term-document matrix. One such technique is to select terms that occur in a minimum number of documents, which can help eliminate rare and unimportant terms. Another widely used approach is term frequency-inverse document frequency (TF-IDF), which assigns a weight to each term based on its frequency in a document and its rarity in the entire corpus (Wu et al., 2008). By filtering terms using these weights, it is possible to retain only the most meaningful and informative terms for further analysis.

The term-document matrix is a vital representation of text data in data mining and information retrieval applications. The application of feature selection techniques, such as TF-IDF filtering, can aid in reducing the size of the matrix and improving the performance of data mining algorithms. This leads to more accurate and meaningful results from the text data analysis.

```
> dtm <- DocumentTermMatrix(docs,)
> dtm
<<DocumentTermMatrix (documents: 12, terms: 7058)>>
Non-/sparse entries: 13518/71178
Sparsity           : 84%
Maximal term length: 383
Weighting          : term frequency (tf)
```

Figure 4.47: Document term matrix (BSI documents)

```
> findFreqTerms(dtm, lowfreq=20)
  [1] "ability"          "acceptance"       "access"
  [4] "according"        "across"           "actions"
  [7] "activities"       "activity"         "additional"
 [10] "address"          "agreed"           "air"
 [13] "analysis"         "ancillary"        "anden"
 [16] "andor"            "applicable"       "application"
 [19] "applications"     "applied"          "applies"
 [22] "appointed"        "appointing"       "appointment"
 [25] "appointments"     "approach"         "appropriate"
 [28] "architecture"     "areas"            "arrangements"
 [31] "aspects"          "assessment"       "assessments"
 [34] "asset"            "assets"           "associated"
 [37] "attributes"       "based"            "basis"
 [40] "benefit"          "built"            "business"
 [43] "capability"       "capacity"         "case"
 [46] "cases"            "chain"            "change"
 [49] "changes"          "check"            "civil"
 [52] "classification"   "clause"           "client"
 [55] "collaborative"    "commercial"       "commissioning"
 [58] "common"           "compile"          "complete"
 [61] "complex"          "complexity"       "compliance"
 [64] "comply"           "components"       "concept"
 [67] "concepts"         "conceptual"       "confirm"
 [70] "consider"         "considered"       "construction"
 [73] "container"        "containers"       "content"
 [76] "context"          "contractor"       "contractors"
 [79] "control"          "controls"         "coordinate"
 [82] "coordination"     "core"             "corresponding"
 [85] "created"          "creating"         "criteria"
```

Figure 4.48: Top 20 words in BSI documents

4.8.2.2 Frequent Terms and Associations

The bar plot shown in Figure 4.49 provides a visual representation of the most frequently used words in the corpus after undergoing some transformations to remove sparse terms. Specifically, the plot displays only the words that appear at least 20 times in the corpus, highlighting the terms that are most commonly used. This type of visualization is useful for gaining a general understanding of the content of the corpus and can provide insights into the themes and topics that are most prevalent. By analysing the bar plot, one can identify the most significant words in the corpus and make inferences about the overall content and structure of the data.

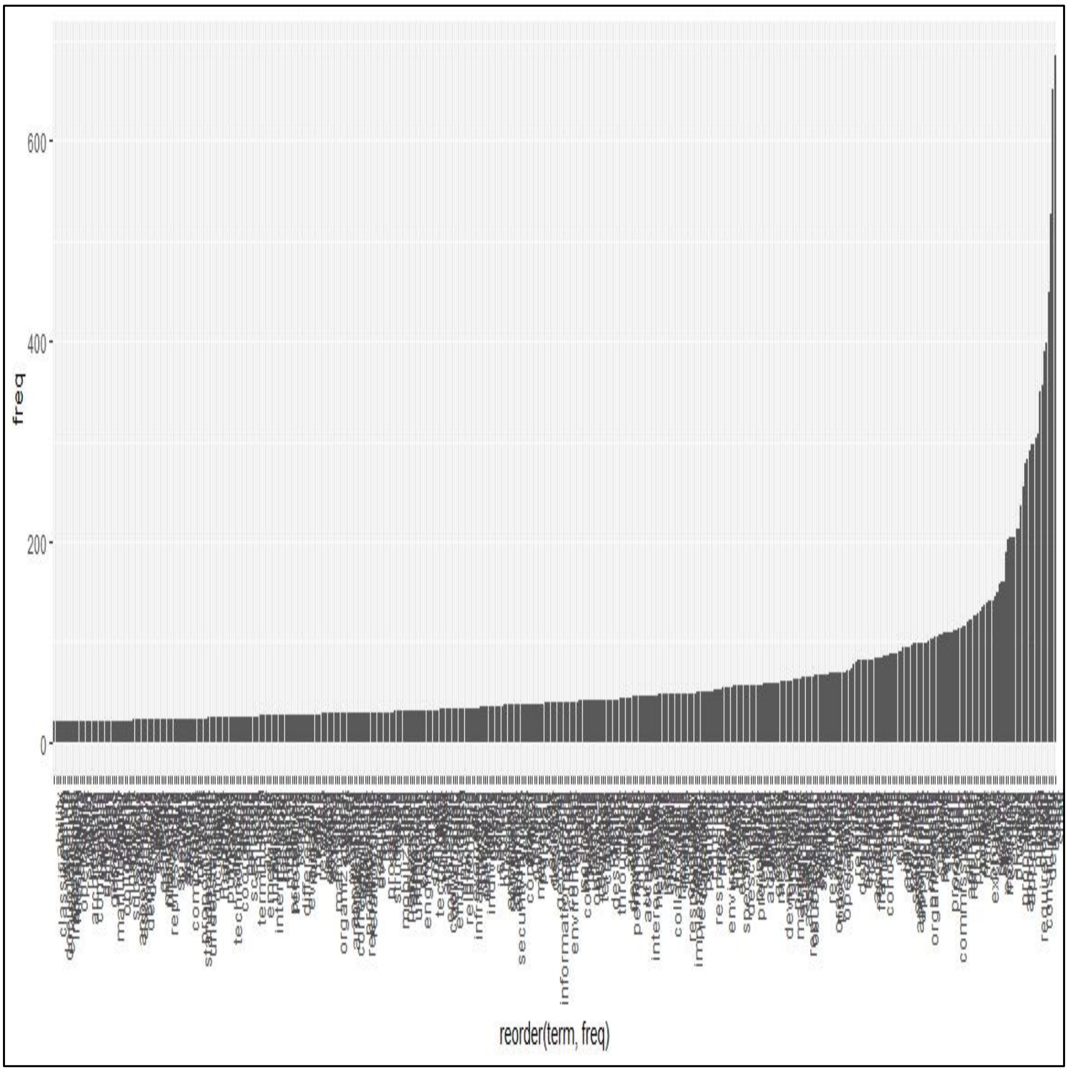


Figure 4.49: A bar plot of Top 20 words in BSI documents

The bar plot in Figure 4.50 and the accompanying Table 4.34 provides a condensed view of the most frequent words in the corpus after limiting the analysis to only 10% of the data. By reducing the size of the corpus, the researcher can gain a more focused understanding of the content of the data and identify the most prominent words and topics. The bar plot and table provide a visual representation of the most frequent words, making it easier to compare the frequency of the different terms and understand their relative importance. This information can be used to guide further analysis and make inferences about the content and structure of the corpus.

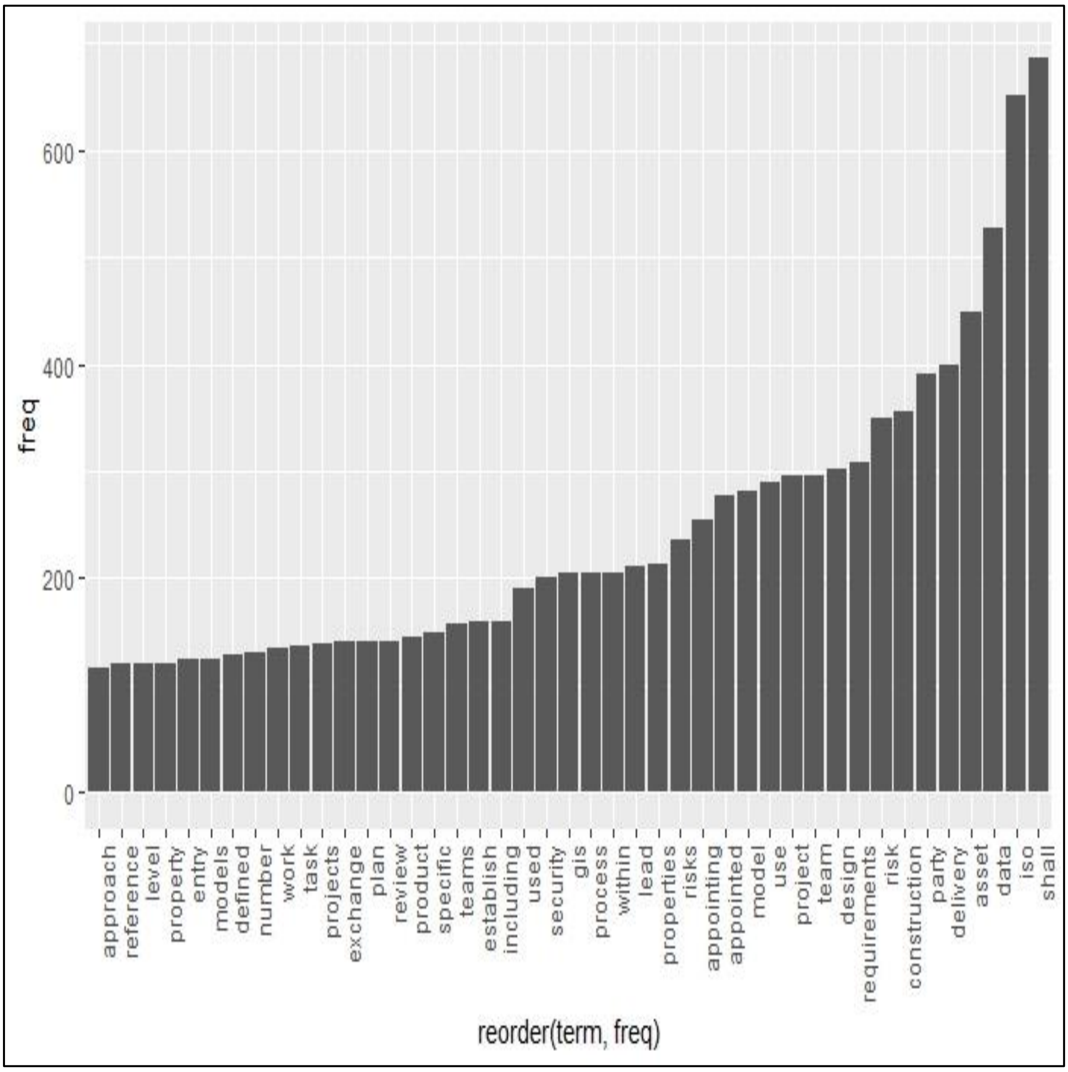


Figure 4.50: A bar plot of frequent words after deleting 10 % of corpus

Table 4.34: Frequencies of top 20 keywords

No.	Keywords	Frequencies
1	Shall	685
2	ISO	651
3	Data	526
4	Asset	449
5	Delivery	399
6	Party	390
7	Construction	355
8	Risk	349
9	Requirements	308
10	Design	303
11	Team	296
12	Project	296
13	Use	290
14	Model	282
15	Appointed	277
16	Risks	235
17	Properties	213
18	Lead	212
19	Within	205
20	GIS	205

The results from the analysis of the most frequent words in the corpus reveal that the word "shall" is the most common term. The repetition of this word highlights the importance placed on the proper archiving of information containers by the appointing party. According to the information presented, upon acceptance of the completed project information model, the appointing party must archive the information containers within the projects common data environment in accordance with the project's information production methods and procedures.

In this process, the appointing party must take into account which information containers will be needed as part of the asset information model for future access requirements and potential future reuse. This emphasis on the proper archiving of information containers highlights the importance of proper information management in the context of the project. The frequency of the word “shall” emphasizes the mandatory nature of these actions, and the significance placed on the proper management of information in the project.

The presence of the words “lead”, “party”, and “appointed” in the frequent word analysis suggests that the lead appointed party has a key role in capturing lessons learned during the project and recording them in a suitable knowledge store for future projects to utilize. This highlights the importance of knowledge transfer and the role that the lead appointed party plays in ensuring that valuable information and lessons are not lost but are captured and made available for future projects.

The frequent use of the word “delivery” indicates that the BSI documents contain important requirements related to the management of information during the delivery phase of built assets. This highlights the importance of proper information management throughout the entire project, including the delivery phase. The requirement for regular review and revision of these requirements until best practice is established underscores the dynamic and evolving nature of information management in the context of built assets.

Finally, the presence of the word “GIS” suggests that information models from both BIM and GIS are becoming increasingly interlinked. This highlights the importance of considering the location and context of each built asset relative to the existing environment. The integration of BIM and GIS information models provides a more comprehensive panorama to understand the built environment, and enables better decision-making and planning for future projects.

This highlights the importance of considering various risks associated with the management of information during the delivery phase of built assets. The assessment of net benefit should take into account the potential consequences, such as financial loss, harm to health and safety, harm to the environment, damage to reputation, and the resources required to capture and maintain information over the asset’s lifecycle. The information models from BIM and GIS must be integrated to support the various practices and requirements within the asset lifecycle, with seamless transitions between both domains from a broader perspective to individual components. Overall, the frequent use of these words in the corpus highlights the crucial role of information management in the delivery phase of built assets.

4.8.2.3 Topic Modelling

In order to obtain the topics discussed in the BSI documents, the researcher applied topic modelling, which is a technique used to extract topics from a set of words. The researcher utilized the Latent LDA model to perform the task. The ‘LDA()’ function was utilized to create a twelve-topic model, which is an appropriate number of topics as there are twelve BSI documents. In other cases, the number of topics may need to be adjusted by trying different values of ‘k’ until the best results are obtained. The resulting topics and the words associated with them are shown in Table 4.35.

Table 4.35: Topic modelling of BSI documents

Topics	Terms	Beta values		Topics	Terms	Beta values
Topic 1	ISO	0.027101		Topic 2	Party	0.041862
	GIS	0.026671			Delivery	0.038494
	Data	0.022137			Appointing	0.028278
	Model	0.017792			Appointed	0.026554
	Schema	0.012360			Shall	0.025827
Topic 3	Party	0.026740		Topic 4	Security	0.044118
	Shall	0.025656			Shall	0.026688
	ISO	0.024636			Organizations	0.019880
	Asset	0.023944			Asset	0.011710
	Appointed	0.017183			Measures	0.010621
Topic 5	Data	0.031516		Topic 6	ISO	0.021378
	Properties	0.021123			Data	0.021097
	ISO	0.020851			Framework	0.010970
	Number	0.015662			Properties	0.010408
	String	0.014956			Shall	0.009845
Topic 7	Risk	0.027134		Topic 8	Asset	0.029266
	Shall	0.026039			Delivery	0.023456
	Design	0.024102			Project	0.020658
	Construction	0.019567			Party	0.014633
	Risks	0.018035			Requirements	0.013342
Topic 9	Party	0.037972		Topic 10	Design	0.030452
	Asset	0.029245			Shall	0.024621
	Shall	0.026887			Construction	0.024208
	Appointed	0.024764			Project	0.023376
	Appointing	0.024764			Asset	0.023043
Topic 11	Data	0.032015		Topic 12	Risk	0.046443
	GIS	0.025020			Design	0.029070
	Mapping	0.024213			Risks	0.019835
	Model	0.023675			Shall	0.019097
	ISO	0.015604			Use	0.014549

These are the topics and the most significant terms that were extracted from the BSI documents using the LDA model. There are 12 topics in total, each with 5 terms. Table 4.35 shows the topic number, the term, and the beta value for each term. The beta value represents the importance of the term in the topic, with higher values indicating higher importance.

Topic 1 seems to be related to ISO standards and data models, with terms such as “ISO”, “GIS”, and “Data” having high beta values. Topic 2 is related to the roles of parties and delivery, with terms such as “Party”, “Delivery”, “Appointing”, and “Appointed”.

Topic 3 is again related to the roles of parties and their responsibilities, with terms such as “Party”, “Shall”, and “Asset”. Topic 4 is related to security and organizations, with terms such as “Security”, “Shall”, and “Organizations”.

Topic 5 is related to data and its properties, with terms such as “Data”, “Properties”, and “ISO”. Topic 6 is related to ISO standards and data framework, with terms such as “ISO”, “Data”, and “Framework”.

Topic 7 is related to risk and its assessment, with terms such as “Risk”, “Shall”, and “Design”. Topic 8 is related to assets and their delivery, with terms such as “Asset”, “Delivery”, and “Project”.

Topic 9 is again related to the roles of parties, with terms such as “Party”, “Asset”, and “Appointed”. Topic 10 is related to design and construction, with terms such as “Design”, “Shall”, and “Construction”.

Topic 11 is related to data and GIS mapping, with terms such as “Data”, “GIS”, and “Mapping”. Topic 12 is related to risk and its management, with terms such as “Risk”, “Design”, and “Risks”.

The organized presentation of the topics identified from the LDA analysis can easily be displayed using the ggplot2 visualization tool (as seen in Figure 4.51).

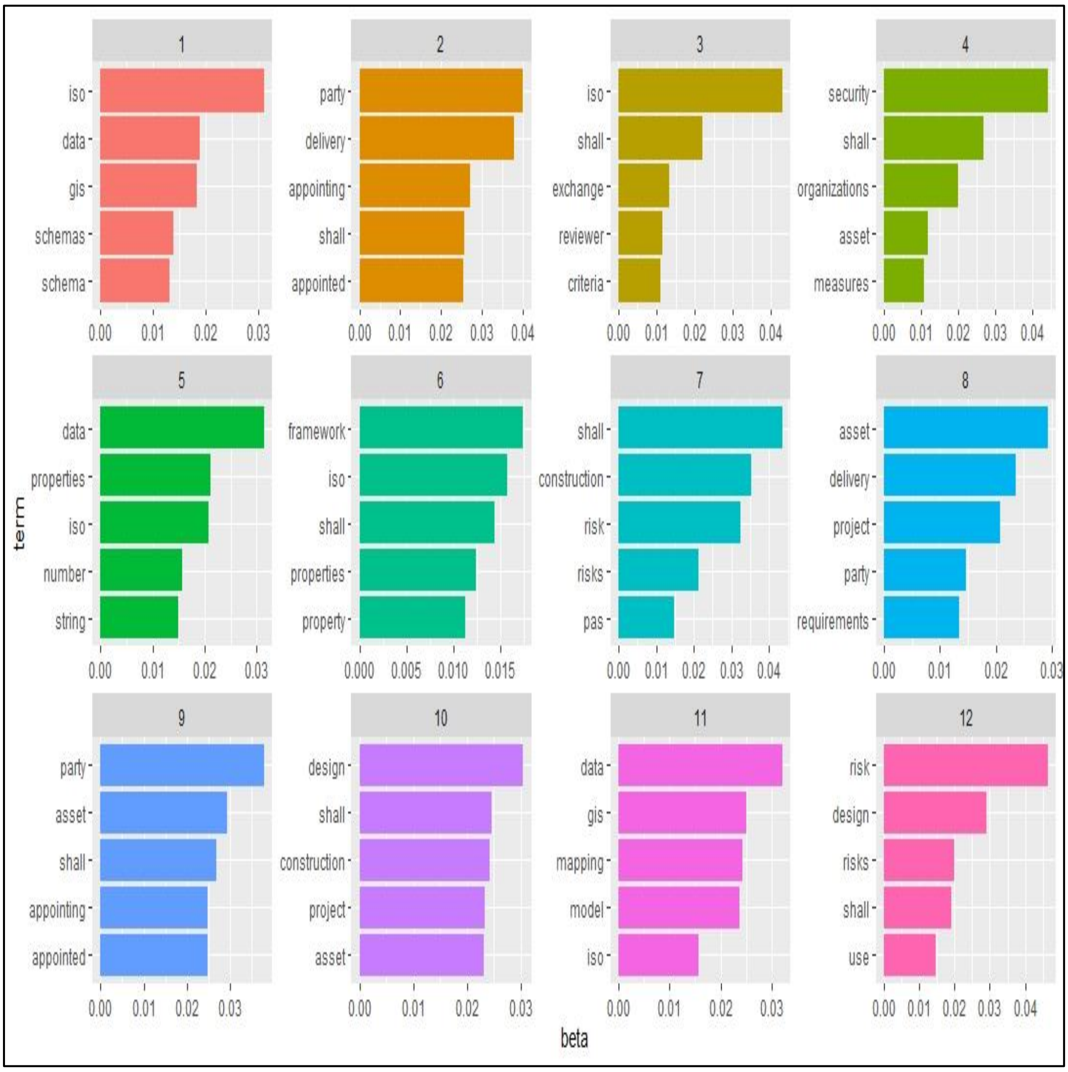


Figure 4.51: A ggplot of BSI topic modelling

4.8.2.4 Word Cloud

This section demonstrates the significance of words by presenting a word cloud visualization (also known as a tag cloud). This type of visualization can be easily produced using the "wordcloud" package (Fellows, 2012). The word cloud in Figure 4.52 effectively reflects the most frequent words present in the term-document matrix created earlier for BSI documents.

The top words in the word cloud, such as "shall", "ISO", "data", "asset", and "delivery", provide insight into the main themes and topics present in the BIM documents. These words suggest that the focus of these documents is on aspects of BIM and how it can be used to improve performance in major projects. Words such as “design”, “model”, “process”, “construction”, “production”, “risk”, and “model” highlight the potential benefits of using BIM in different phases of a project’s lifecycle, including design, build, and operation.

Additionally, the results of the word cloud lead the researcher to conclude that the use of BIM can mitigate risks and provide a better model for simulation and risk management by combining BIM and GIS. The frequent use of words such as “asset”, “hazards”, and “client” also suggest that the creation of a client information model is an important aspect of BIM.

Furthermore, the presence of words such as “IFC” and “security” implies that the ownership and protection of BIM data is a topic of concern. This highlights the importance of ensuring that the data generated and used during a BIM project is secure and can be easily shared among stakeholders.

Figure 4.52: Word cloud of 20 frequent words (BSI documents)

In an attempt to further analyse the results, the researcher applied hierarchical clustering to find clusters of words. Hierarchical clustering is a powerful tool, but it may not always produce the most accurate results; it is thus important to consider the limitations and assumptions inherent in the use of this method. Hierarchical clustering results may be influenced by the choice of similarity measure, distance metric, and linkage method used. Additionally, clustering results can be sensitive to the number of clusters specified, and the interpretation of the results may be subjective.

While the word cloud visualization and the use of hierarchical clustering are useful tools for exploring the main themes and topics present in the BSI documents, it is important to critically evaluate the results and consider the limitations and assumptions inherent in the methods used.

4.8.2.5 Clustering Words

For the purpose of this experiment, the researcher employed the code previously demonstrated in section 4.7.3.5 to identify clustering words. Figure 4.53 displays the clustering words extracted from BSI documents after eliminating sparse terms and reducing the words to 0.15%. The researcher then divided these clustering words into 12 groups, as illustrated in Figure 4.54.

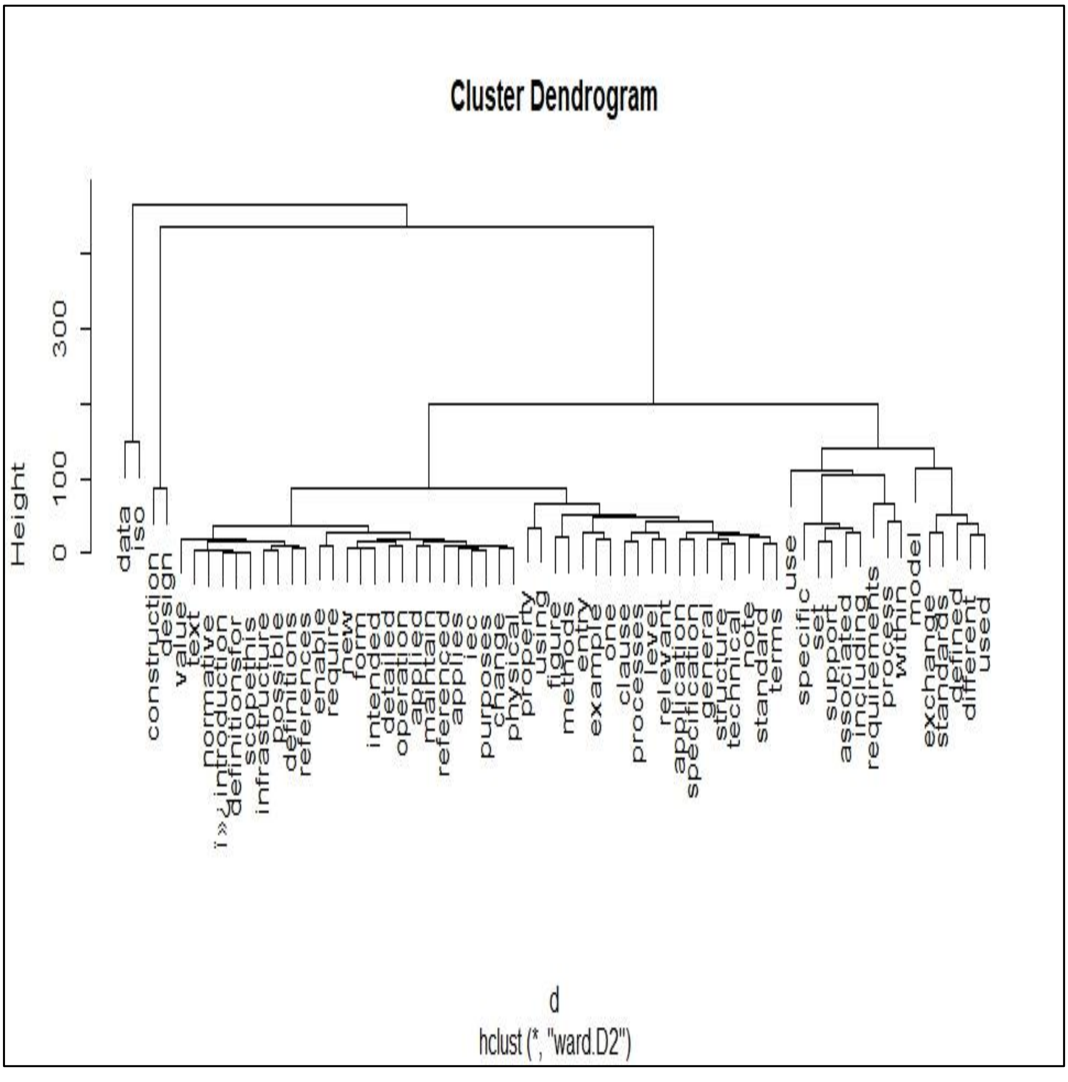


Figure 4.53: The clustering words of BSI documents

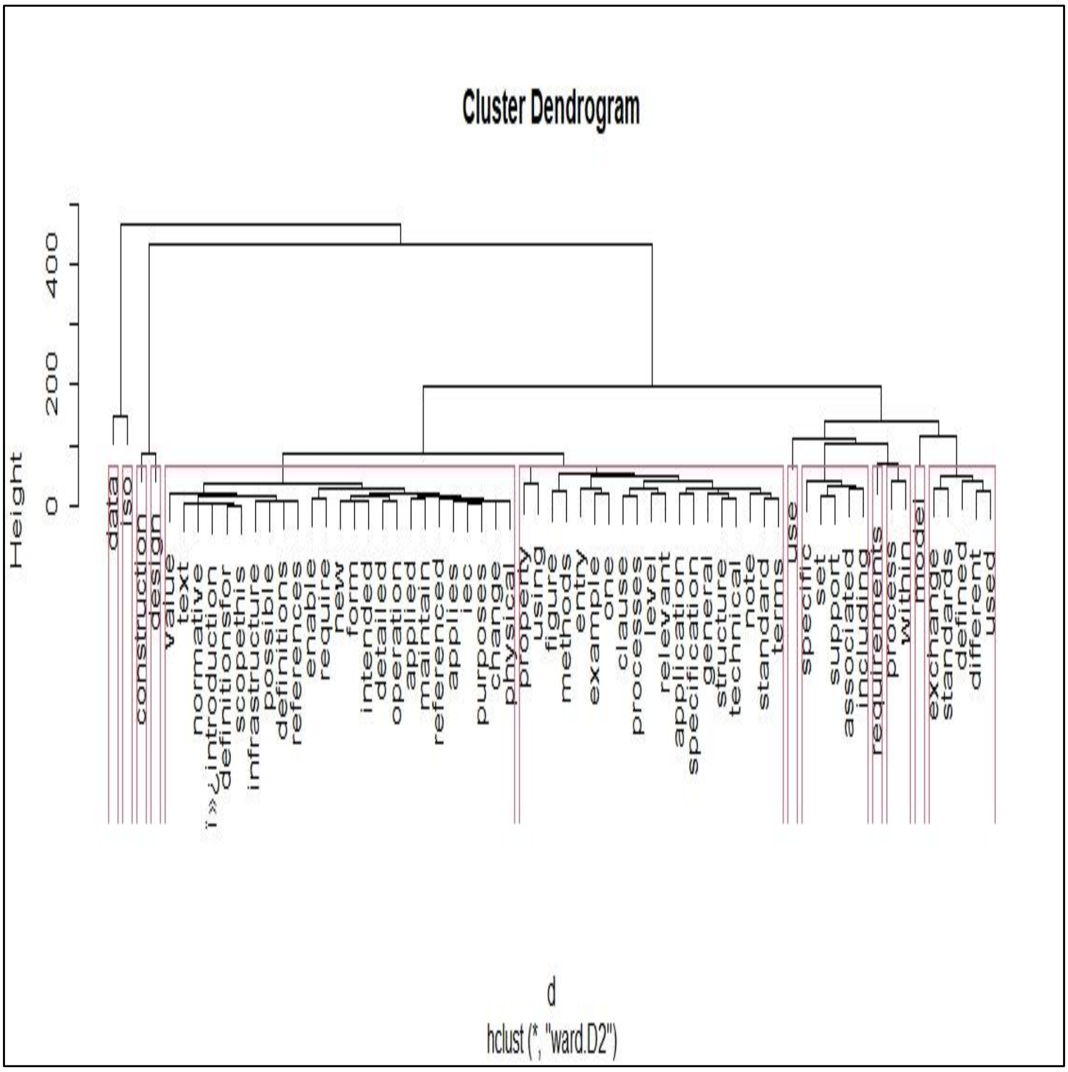


Figure 4.54: 12 groups of the clustering words

Figure 4.54 provides a visual representation of the word clusters obtained from the k-means clustering algorithm. It can be observed that certain words, such as “ISO”, “requirements”, “entry”, etc., are grouped together in one cluster. Another cluster is composed of words such as “ISO”, “process”, “model”, etc. By using k-means algorithms to cluster the documents, the researcher aims to gain deeper insights into the relationships between the words in the BSI documents.

4.8.2.6 Clustering Documents with the K-Means Algorithm

The researcher applied the k-means clustering algorithm to the term-document matrix, which was transformed into a document-term matrix. The algorithm grouped the documents into twelve clusters using the ‘k-means()’ function. The researcher then examined the popular words in each cluster, as well as the cluster centres. To make it easier to understand the themes of each cluster, the researcher identified the top five words in each cluster. The code used for this analysis is shown in Figure 4.55, and Table 4.36 displays the top five words in each cluster.

```
> kmeansResult <- kmeans(m3, k)
> round(kmeansResult$centers, digits = 5)
   application applied applies associated change clause construction  data defined
1            2       8       1        4.0      4      1         20.0   6.0    15.0
2            4       9       2        9.0      3      5          3.0   6.0     2.0
3            1       6       1        2.0      7     12          0.0  11.0     2.0
4            1       5       1        9.0      2      3          7.0  18.0     8.0
5            5       1       1        1.0      1      3          3.0  55.0    31.0
6            7       1       1        0.5      2      5          3.5   5.5     1.5
7            3       1       1        4.0      0     11         71.0 222.0    14.0
8           23       2       4        4.0      7      8         16.0 115.0    38.0
9            6      11       8       55.0      7     21        209.0  18.0    10.0
10           4       6       1        5.0      4     12          1.0   0.0     4.0
11           0       1       1        2.0      1      2         18.0  64.0     2.0
   definitions definitionsfor design detailed different enable entry example
1          1.0              1   14.0       12      19.0    4.0  30.0    21.0
2          2.0              1    1.0        5       7.0    5.0   4.0    11.0
3          3.0              1    1.0        3       1.0    1.0  12.0     3.0
4          2.0              1    2.0        5       1.0    1.0   3.0     1.0
5          1.0              1    1.0        4       1.0    1.0   7.0     4.0
6          2.5              1    4.5        1       2.5    0.5   1.5     1.5
7          2.0              1    4.0        3      28.0   10.0  22.0    16.0
8          9.0              1    4.0        3      38.0    4.0   9.0     3.0
9          2.0              0  259.0        8       8.0   20.0   5.0    13.0
10         1.0              1    3.0        1       1.0    1.0  10.0     3.0
11         1.0              1    5.0        0       6.0    2.0  20.0     3.0
   exchange figure form general ï»¿introduction iec including infrastructure
1        18   11.0  7.0     4.0             1.0   4        17              2
2        10   28.0  1.0     1.0             1.0   2        22              4
3        23   15.0  0.0     1.0             1.0   2        14              2
4        20   12.0  2.0     2.0             1.0   2        10              7
5         1    5.0  3.0     4.0             1.0   2         9              3
6         2    0.5  2.5     2.5             0.5   2         4              1
7         2   11.0  2.0    18.0             1.0   2         6              1
8        51    7.0  4.0    15.0             1.0   2         8              5
9         6   13.0  5.0    13.0             1.0   8        46              3
10        0    3.0  2.0     4.0             1.0   2        17              4
11        6    3.0  2.0     1.0             1.0   2         3              1
   intended iso level maintain methods model  new normative note  one operation
1       5.0  38    11      3.0     4.0    19 10.0         2   13 26.0        10
2       2.0  63    10     10.0    25.0    40  5.0         1   10  2.0         8
3       2.0  74     1      1.0     5.0     7  1.0         1   16  1.0         2
4       3.0  13    11      2.0    26.0    40  2.0         1    3  1.0         1
5       0.0  16     3      1.0     3.0    82  0.0         1    2  2.0         9
6       4.5  17     2      1.5     1.5     5  4.5         2    2  4.5         1
```

Figure 4.55: K-means results

Table 4.36: Top five words in every cluster

Clusters	Strings of Keywords				
Cluster 1	Requirements	ISO	Entry	One	Within
Cluster 2	ISO	Process	Within	Model	Requirements
Cluster 3	ISO	Exchange	Using	Note	Figure
Cluster 4	Within	Requirements	Model	Methods	Exchange
Cluster 5	Model	Data	Defined	Property	Use
Cluster 6	ISO	Property	Requirements	Application	Specific
Cluster 7	Data	ISO	Construction	Property	Use
Cluster 8	ISO	Data	Model	Exchange	Standards
Cluster 9	Design	Construction	Use	Requirements	Associated
Cluster 10	Requirements	Use	ISO	Within	Including
Cluster 11	Data	ISO	Entry	Construction	Used

Table 4.28 presents the results of a k-means clustering analysis of certain words found in the BSI documents. The analysis grouped the words into 11 clusters, where each cluster represents a group of words that are related to each other in meaning or frequency of use.

Cluster 1 includes words such as “Requirements”, “ISO”, “Entry”, and “Within”. This cluster likely represents a group of words related to setting standards, guidelines, and requirements for processes or procedures.

Cluster 2 includes words such as “ISO”, “Process”, “Within”, “Model”, and “Requirements”. This cluster likely represents a group of words related to the development, implementation, and management of processes and models within an organizational context.

Cluster 3 includes words such as “ISO”, “Exchange”, “Using”, “Note”, and “Figure”. This cluster likely represents a group of words related to exchanging information, data, and standards through various means, such as notes and figures.

Cluster 4 includes words such as “Within”, “Requirements”, “Model”, “Methods”, and “Exchange”. This cluster likely represents a group of words related to establishing requirements, models, and methods within an organizational context and exchanging information between stakeholders.

Cluster 5 includes words such as “Model”, “Data”, “Defined”, “Property”, and “Use”. This cluster likely represents a group of words related to developing and using models, data, and defined properties.

Cluster 6 includes words such as “ISO”, “Property”, “Requirements”, “Application”, and “Specific”. This cluster likely represents a group of words related to establishing requirements, properties, and specific applications within an organizational context.

Cluster 7 includes words such as “Data”, “ISO”, “Construction”, “Property”, and “Use”. This cluster likely represents a group of words related to the collection, use, and construction of data and properties within an organizational context.

Cluster 8 includes words such as “ISO”, “Data”, “Model”, “Exchange”, and “Standards”. This cluster likely represents a group of words related to the exchange of data, models, and standards within an organizational context.

Cluster 9 includes words such as “Design”, “Construction”, “Use”, “Requirements”, and “Associated”. This cluster likely represents a group of words related to the design, construction, use, and association of requirements within an organizational context.

Cluster 10 includes words such as “Requirements”, “Use”, “ISO”, “Within”, and “Including”. This cluster likely represents a group of words related to establishing requirements, use, and inclusion within an organizational context.

Cluster 11 includes words such as “Data”, “ISO”, “Entry”, “Construction”, and “Used”. This cluster likely represents a group of words related to the collection, entry, and use of data and construction within an organizational context.

The relationship between these word clusters and BIM in major construction project delivery can be understood by considering the role of BIM in such projects. The word clusters listed in Table 4.28 may represent various aspects of BIM-based project delivery. For example, “ISO” and “standards” in Clusters 1, 2, 8, and 10 may refer to the standardized processes and protocols that are used in BIM-based project delivery. “Requirements”, “entry”, “model”, and “methods” in Clusters 1, 2, 4, and 6 may refer to the essential components of BIM-based project delivery, such as the requirements of the project, the data entry process, the BIM model, and the methods used to manage the project. “Data”, “property”, “use”, and “construction” in Clusters 5, 7, 9, and 11 may refer to the data management, data exchange, and the use of BIM in construction processes.

Overall, the word clusters in Table 4.28 can be used to identify key areas of BIM-based project delivery and to understand how different aspects of the delivery process are interrelated. By understanding these relationships, it becomes possible to identify areas where performance can be enhanced in major construction project delivery. For example, if the requirements for a project are not well defined or are not being met, it may be possible to improve performance by optimizing the requirements management processes. If the ISO standards are not being adhered to, it may be necessary to revise the project’s processes to ensure compliance. Additionally, by understanding the relationships between data, models, and construction, it becomes possible to improve data management and construction processes, leading to enhanced project delivery. The word clusters provide a way to identify areas of the project delivery process that are critical to performance, making it possible to focus improvement efforts in the right areas.

4.8.3 Lessons Learned from Experiment 4

In the process of analysing the BSI documents, several techniques were employed to gain insights and extract meaningful information. One of the initial steps involved building a term-document matrix, as depicted in Figure 4.47, which consisted of 7,058 terms and 12 documents. The matrix exhibited a high degree of sparsity, with 84% of its entries being zero. To effectively perform data mining tasks on such a large matrix, feature selection techniques were employed to reduce its size. Two common techniques used for feature selection were discussed. The first technique involved selecting terms that occur in a minimum number of documents to eliminate rare and unimportant terms. The second technique, TF-IDF, assigned a weight to each term based on its frequency in a document and rarity in the entire corpus. By filtering terms using these weights, only the most meaningful and informative terms were retained for further analysis.

The term-document matrix served as a foundation for various data mining tasks, including clustering, classification, and association analysis. By reducing the size of the matrix through feature selection, the performance of data mining algorithms can be improved, leading to more accurate and meaningful results in text data analysis. The most frequently used terms in the corpus were analysed and presented in Figure 4.48. This analysis provided insights into the content and structure of the data. A bar plot of the most frequent words, as shown in Figure 4.59, further visualized the commonly used terms in the corpus, highlighting the most prevalent themes and topics.

The bar plot and accompanying table, as seen in Figure 4.50 and Table 4.34, condensed the analysis by considering only 10% of the data. This reduction allowed for a more focused understanding of the content and identification of the most prominent words and topics. The table shows the frequency of the 20 most frequent words, providing a comparative view of their importance. The analysis of the most frequent words revealed several significant findings. The word “shall” was the most common term, emphasizing the importance of proper archiving of information containers by the appointing party. The presence of words like “lead”, “party”, and “appointed” highlighted the role of the lead appointed party in capturing and storing valuable knowledge for future projects. The frequent use of the word “delivery” indicated important requirements related to information management during the delivery phase of built assets. Additionally, the word “GIS” suggested the increasing integration of BIM and GIS models, emphasizing the consideration of location and context in asset management.

Topic modelling was employed to extract the topics discussed in the BSI documents. The LDA model, utilizing the ‘LDA()’ function, was applied to create a twelve-topic model, aligning with the number of documents. The resulting topics and associated terms are presented in Table 4.35, offering insights into the main themes and subjects covered in the documents. Each topic was represented by a set of significant terms, with higher beta values indicating greater importance. To visually represent the topics extracted, a ggplot visualization, as shown in Figure 4.51, was utilized. This presentation facilitated a clearer understanding of the distribution and relationships between the topics.

A word cloud, displayed in Figure 4.52, was created to further illustrate the significance of words in the term-document matrix. The word cloud showcased the most frequent words in the BSI documents and provided insights into the main themes and topics. Various words, such as “shall”, “ISO”, “data”, “asset”, and “delivery” reflected the focus on BIM and its application in project performance improvement. Other words highlighted different phases of a project’s lifecycle, including design, construction, and operation.

The clustering results shown in Figure 4.54 provide a visual representation of the word clusters obtained from the k-means clustering algorithm. Each cluster is assigned a unique colour, and the words within each cluster are grouped together. By examining the clusters, patterns and relationships between words can be identified. For example, in Cluster 1, words such as “ISO”, “requirements”, and “entry” are grouped together, suggesting a relationship between ISO standards and requirements related to data entry. Cluster 2 contains words such as “ISO”, “process”, and “model”, indicating a connection between ISO standards and the process of modelling. These clusters provide insights into the main themes and topics present in the BSI documents and help uncover the relationships between different concepts.

By using clustering techniques, the researcher can identify common patterns, themes, and topics within the BSI documents. This information can be valuable for various TM tasks, such as topic analysis, content categorization, and information retrieval. Clustering allows for a deeper understanding of the textual data and can aid in the development of more targeted strategies for analysing and extracting knowledge from the documents. It is important to note that the effectiveness of clustering depends on various factors, including the choice of clustering algorithm, similarity measure, and the quality and representativeness of the data. Additionally, the interpretation of the clustering results requires domain knowledge and careful analysis to derive meaningful insights.

Overall, the application of clustering techniques to the BSI documents enables the researcher to uncover patterns, relationships, and themes within the textual data, facilitating a more comprehensive understanding of the information contained in the documents and supporting further TM and analysis tasks.

4.9 Experiment 5: TS 5 BIM Policy at Companies TM Process

4.9.1 Pre-processing of the Raw Textual Data

The researcher has utilized the results from the analysis of 24 BIM policy papers, which were obtained and described in accordance with the criteria outlined in section 4.6.2.3. To avoid repetition, the text cleaning process was repeated. The researcher then tabulated the company names and coverage of each policy, which is displayed in Table 4.26. The determination of the coverage of each policy was based on the author's subjective evaluation through reading each policy paper individually. However, to mitigate any potential biases, the researcher sought to identify the most commonly covered topics in the policies of the various companies and compared these results with the data presented in the existing table. The researcher performed similar text analysis tasks as in previous sections, and has excluded the words “policy”, “building”, “information”, “modelling”, “management”, and “BIM” from the corpus.

4.9.2 New Knowledge

4.9.2.1 Building A Term-Document Matrix

The term-document matrix, as presented in Figure 4.56, is comprised of 3,037 terms and 24 documents. The matrix is characterized by its high degree of sparsity, with 92% of entries being zero. The researcher analysed the most frequently used terms within the matrix (refer to the code in Figure 4.57). The term-document matrix presents ample opportunities for data mining tasks such as clustering, classification, and association analysis. When faced with an excessive number of terms, it is possible to reduce the size of the term-document matrix by selecting terms that appear in a minimum number of documents or by applying a TF-IDF filter (Wu et al., 2008).

```
> dtm <- DocumentTermMatrix(docs,)
> dtm
<<DocumentTermMatrix (documents: 24, terms: 3034)>>
Non-/sparse entries: 6012/66804
Sparsity           : 92%
Maximal term length: 119
Weighting          : term frequency (tf)
```

Figure 4.56: Document term matrix (BIM policy at companies' papers)

```
> findFreqTerms(dtm, lowfreq=20)
 [1] "achieve"        "across"         "adoption"       "benefits"
 [5] "best"           "capability"     "client"         "clients"
 [9] "collaborative"  "committed"      "common"         "company"
[13] "construction"   "coordination"   "data"           "delivery"
[17] "design"         "development"    "digital"        "efficient"
[21] "engineering"    "ensure"         "environment"    "government"
[25] "group"          "ifc"            "implementation" "including"
[29] "industry"       "infrastructure" "level"          "ltd"
[33] "model"          "models"         "new"            "pas"
[37] "plan"           "practice"       "process"        "processes"
[41] "project"        "projects"       "public"         "queensland"
[45] "requirements"   "review"         "software"       "staff"
[49] "standards"      "strategy"       "support"        "team"
[53] "technology"     "time"           "training"       "using"
[57] "work"           "working"
```

Figure 4.57: Code for finding frequent terms

4.9.2.2 Frequent Terms and Associations

The researcher has created a bar plot to represent the frequently used words in the corpus, which were filtered to include only those that appeared at least 20 times. The plot, which is displayed in Figure 4.58, provides a visual representation of the most common terms used in the BIM policy documents from various companies. Additionally, Table 4.27 lists the top 20 most frequently used words in these documents.

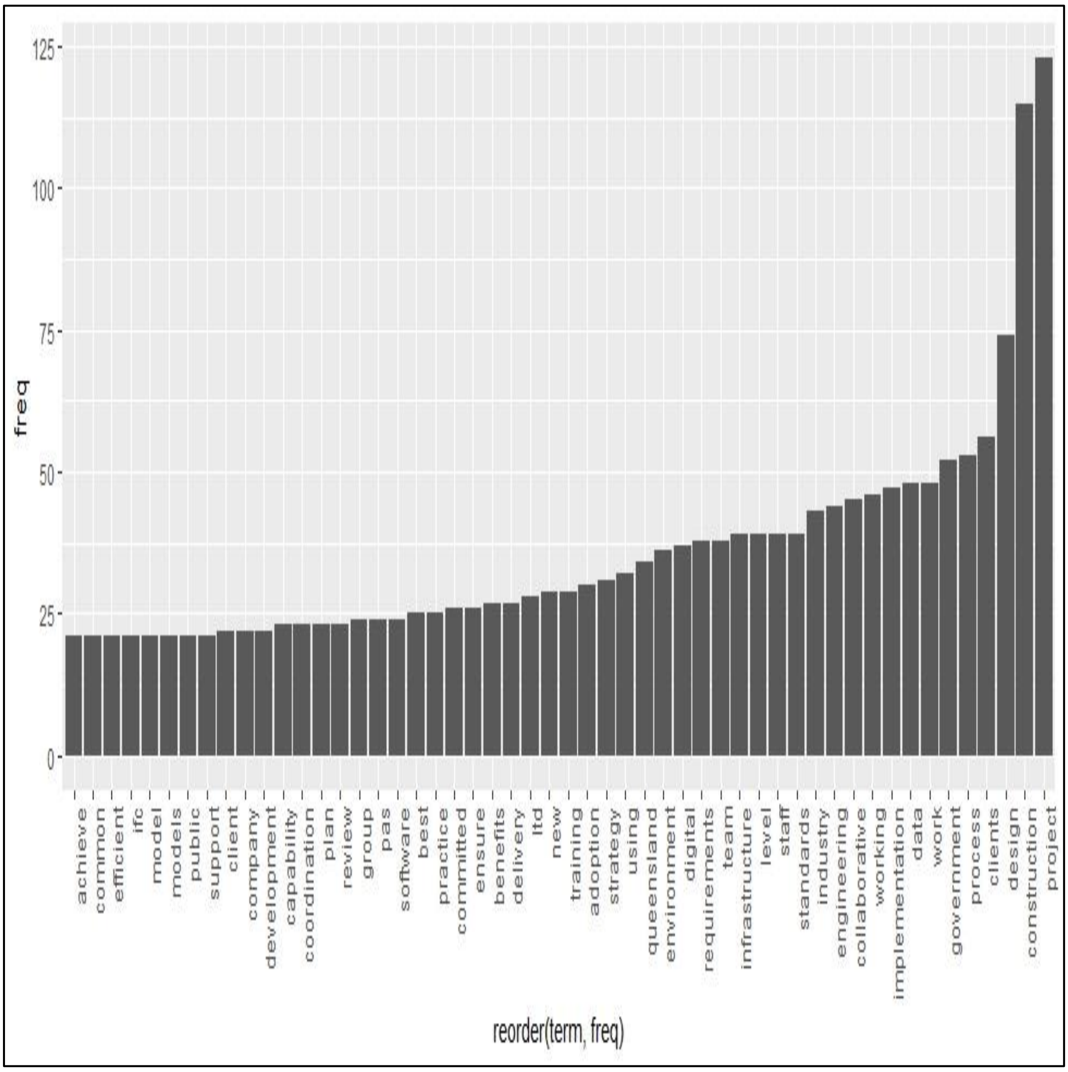


Figure 4.58: A bar plot of frequently used words in BIM TS 5

Table 4.37: Frequencies of top 20 keywords in BIM TS 5

No.	Keywords	Frequencies
1	Project	123
2	Construction	115
3	Design	74
4	Clients	56
5	Process	53
6	Government	52
7	Work	48
8	Data	48
9	Implementation	47
10	Working	46
11	Collaborative	45
12	Engineering	44
13	Industry	43
14	Standards	39
15	Staff	39
16	Level	39
17	Infrastructure	39
18	Team	38
19	Requirements	38
20	Digital	37

The analysis of the most frequently used words in the BIM policy papers of different companies reveals important insights into the coverage and scope of these policies. The most frequent word, “project”, can be considered a general term and can therefore disregarded. The other most frequent words, such as “construction”, “design”, “client”, and “process”, suggest that the BIM policies cover all phases of the project lifecycle, from design and planning to construction and delivery.

The frequency of the word “government” indicates that these companies may be following guidelines and regulations established by government entities. This is a critical aspect of BIM-based project delivery, as it helps to ensure that projects are executed in a way that meets the necessary standards and regulations.

The word “collaborative” highlights the importance of collaboration in BIM as a process and platform. BIM has been widely adopted in the construction industry due to its ability to facilitate collaboration among all stakeholders involved in a project, including architects, engineers, contractors, and owners.

Words such as “engineering”, “standards”, and “industry” suggest that the companies who have BIM policies are following international standards, such as ISO 14000 certification, which is an international Environmental Management Guideline (EMG) for product lifecycle and sustainability standard for business, industry, and engineering design.

Finally, the word “infrastructure” with a frequency of 39 implies that related companies cover major infrastructure projects within their BIM policies. This highlights the significance of BIM in delivering large-scale, complex projects that have a significant impact on the built environment.

The analysis of the most frequently used words in BIM policy papers can provide valuable insights into the coverage and scope of these policies and help to identify the key areas that companies focus on when implementing BIM in their project delivery processes.

4.9.2.3 Association and Topic Modelling

The function findAssocs() was employed to analyse the terms associated with certain words in the corpus. In particular, the researcher has focused on identifying terms that are correlated with performance aspects such as time, cost, and quality, with a correlation coefficient of at least 0.85. The results of the analysis indicate that only one term, “effort”, is strongly associated with the term “performance” in the BSI policy documents. This information can be seen in Figure 4.59. The word “effort” is associated with “performance” as they both refer to the level of energy, time, and resources that are put into a task or project. Effort can be considered as a measure of the inputs used in the process of achieving a certain performance outcome. In the context of BIM-based project delivery, effort can refer to the level of work and resources that are required to meet performance goals in terms of time, cost, and quality. The presence of the word “effort” in the documents analysed suggests that companies are aware of the importance of ensuring that the right level of resources are allocated to BIM projects in order to achieve desired performance outcomes.

```
> fn_wordAssociations = function(dtm, aterm)
+ {
+   findAssocs(tdm, aterm, 0.85)
+ }
> assocWords = 'performance'
> assocWords.association = lapply(assocWords, fn_wordAssociations, dtm = dtm)
> print(assocWords.association)
[[1]]
[[1]]$performance
effort
  0.91
```

Figure 4.59: Words associated with the word 'performance'

In contrast, no words were found to be associated with the terms “cost” and “quality”, suggesting that these policy documents do not address these aspects. However, multiple words were found to be associated with the term “time”. These include words such as “productivity”, “rework”, and “errors”, which indicate that these companies have implemented BIM policies to effectively use them in order to minimize the occurrence of rework problems caused by design errors, improve the project delivery rate, enhance productivity, and meet project deadlines (as shown in Figure 4.60). The inclusion of these terms in the policy documents highlights the importance that these companies place on time management, as well as the recognition of the role that BIM can play in avoiding costly design errors and rework. Overall, the words “productivity”, “rework”, and “errors” indicate a focus on improving efficiency and reducing waste in the project delivery process through the use of BIM.

```
> assocWords = c('cost', 'time', 'quality')
> assocWords.association = lapply(assocWords, fn_wordAssociations, dtm = dtm)
> print(assocWords.association)
[[1]]
[[1]]$cost
numeric(0)


[[2]]
[[2]]$time
          aec         back      example         main     overcome productivity
         0.91         0.91         0.91         0.91         0.91         0.91
   resistance       rework      several      similar         term        users
         0.91         0.91         0.91         0.91         0.91         0.91
        value      general    essential        years       create      details
         0.91         0.90         0.89         0.89         0.88         0.87
   determined   eventually     flexible         long          one   overcoming
         0.87         0.87         0.87         0.87         0.87         0.87
 particularly     strongly  contractors    different       models          set
         0.87         0.87         0.86         0.86         0.86         0.86
       errors         task        three
         0.85         0.85         0.85


[[3]]
[[3]]$quality
numeric(0)
```

Figure 4.60: Words associated with cost, time, and quality

The next step after finding the associations between words is to determine what topics are most commonly addressed by the companies in their BIM policies. This information can be obtained through topic modelling, which involves identifying latent themes in a corpus of text data. By comparing the results of the topic modelling with the existing table presented in section 4.8.2.3, the researcher can gain insights into what aspects of BIM are most commonly addressed by these companies in their policies.

4.9.2.4 Topic Modelling

The results of the topic modelling show that construction and design are the most frequently discussed topics in the BIM policies. These findings are in line with the results of the previous analyses of the most frequent words, which also showed construction and design to be among the most frequently mentioned words. This highlights the importance that these companies place on ensuring that BIM is effectively integrated into the construction and design phases of a project.

Topic 1 contains the word “project”, which highlights the central role that projects play in the BIM policies. The presence of the word “impact” in this topic may indicate that these companies are aware of the potential impact that BIM can have on various aspects of a project, such as time, cost, and quality.

Topic 2, which consists of words such as “queensland”, “government”, “infrastructure”, “principles”, “planning”, and “department”, highlights the involvement of the government in BIM policies. This may indicate that these companies are closely following the guidelines and regulations set by the government in regard to the use of BIM in their projects.

The presence of words such as “infrastructure” and “planning” in Topic 2 highlights the potential benefits that BIM can bring to major infrastructure projects, such as improved planning and increased efficiency. The other topics discussed in the BIM policies also offer insights into the various aspects that these companies consider when implementing BIM in their projects.

Overall, the results of the topic modelling help to provide a more comprehensive view of the topics covered in the BIM policies of these companies and highlights the importance that these companies place on effective implementation of BIM in their projects. Other topics are illustrated in Figure 4.61.

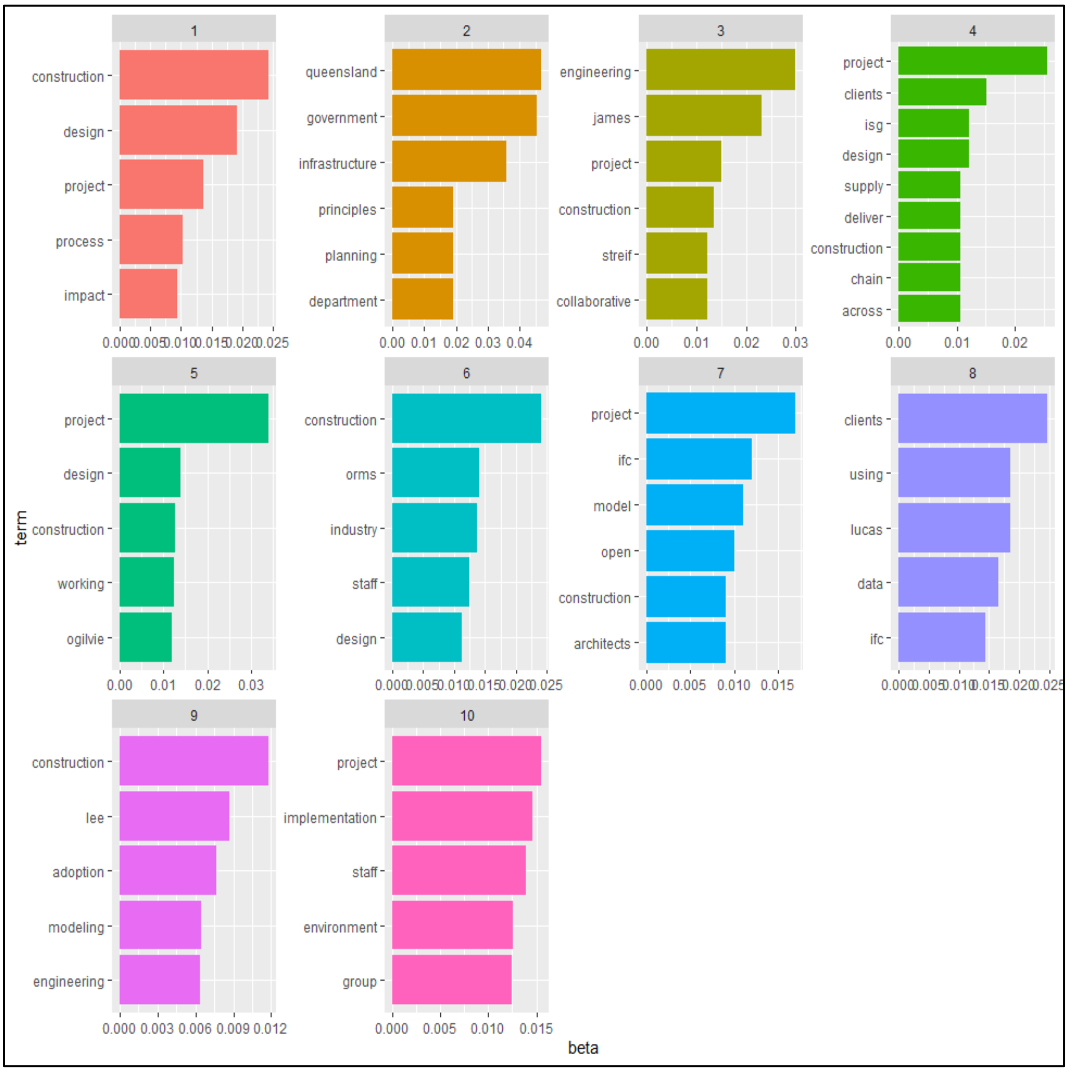


Figure 4.61: Topic modelling of BIM policy papers at companies

4.9.2.5 Word Cloud

The analysis of the BIM policy documents has shown that these documents include information related to BIM. This has been validated through the word cloud in Figure 4.62, where the researcher has identified a set of important words such as “implementation”, “design”, “client”, “requirements”, “digital”, “team working”, “software development”, and “coordination”. Another set of frequent words includes “staff training”, “achieve efficiency”, “IFC”, “modelling”, “planning”, and “supply chain”. The presence of words related to “people”, “processes”, and “products” highlights the focus on digital support and software development in these policies.

Moreover, the results from the topic modelling analysis further reinforce the insights obtained from the word cloud, as the top 10 topics discussed in the BIM policies include “construction”, “design”, “project”, and “impact” in Topic 1, and “queensland”, “government”, “infrastructure”, principles”, “planning”, and “department” in Topic 2. These topics demonstrate a focus on various aspects of BIM, including design, construction, and government regulations, as well as the impact of these policies on the infrastructure and project planning processes. In the next stage of analysis, the researcher used hierarchical clustering to group similar words and extract further insights from the BIM policy documents.

Figure 4.62: Word cloud of BIM policy documents

4.9.2.6 Clustering Words

The researcher utilized the ‘hclust()’ function to cluster the relevant terms and produced a dendrogram which was divided into 10 clusters, as depicted in Figure 4.63. Subsequently, the researcher divided these clustered terms into 10 distinct groups as depicted in Figure 4.64.

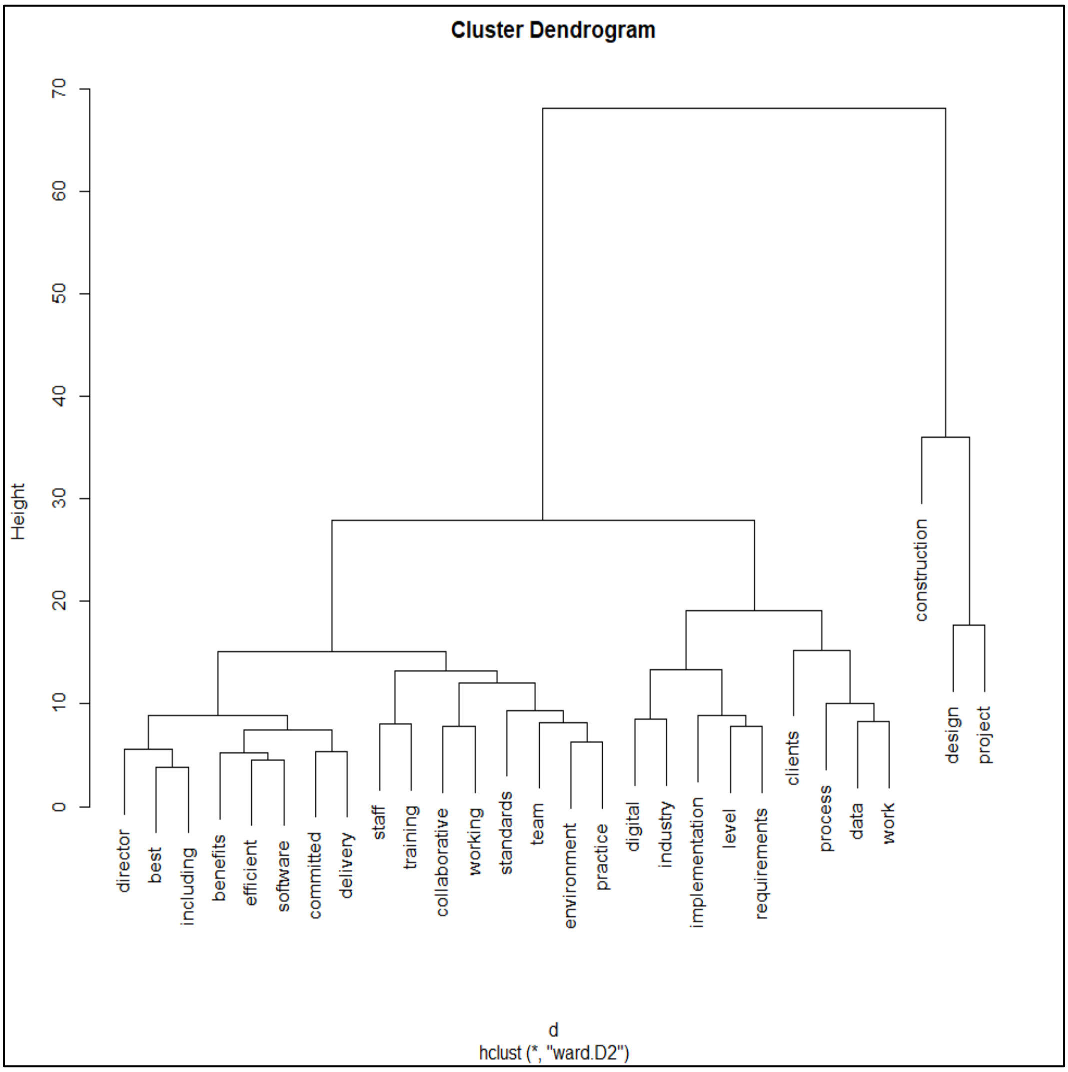


Figure 4.63: Cluster analysis dendrogram

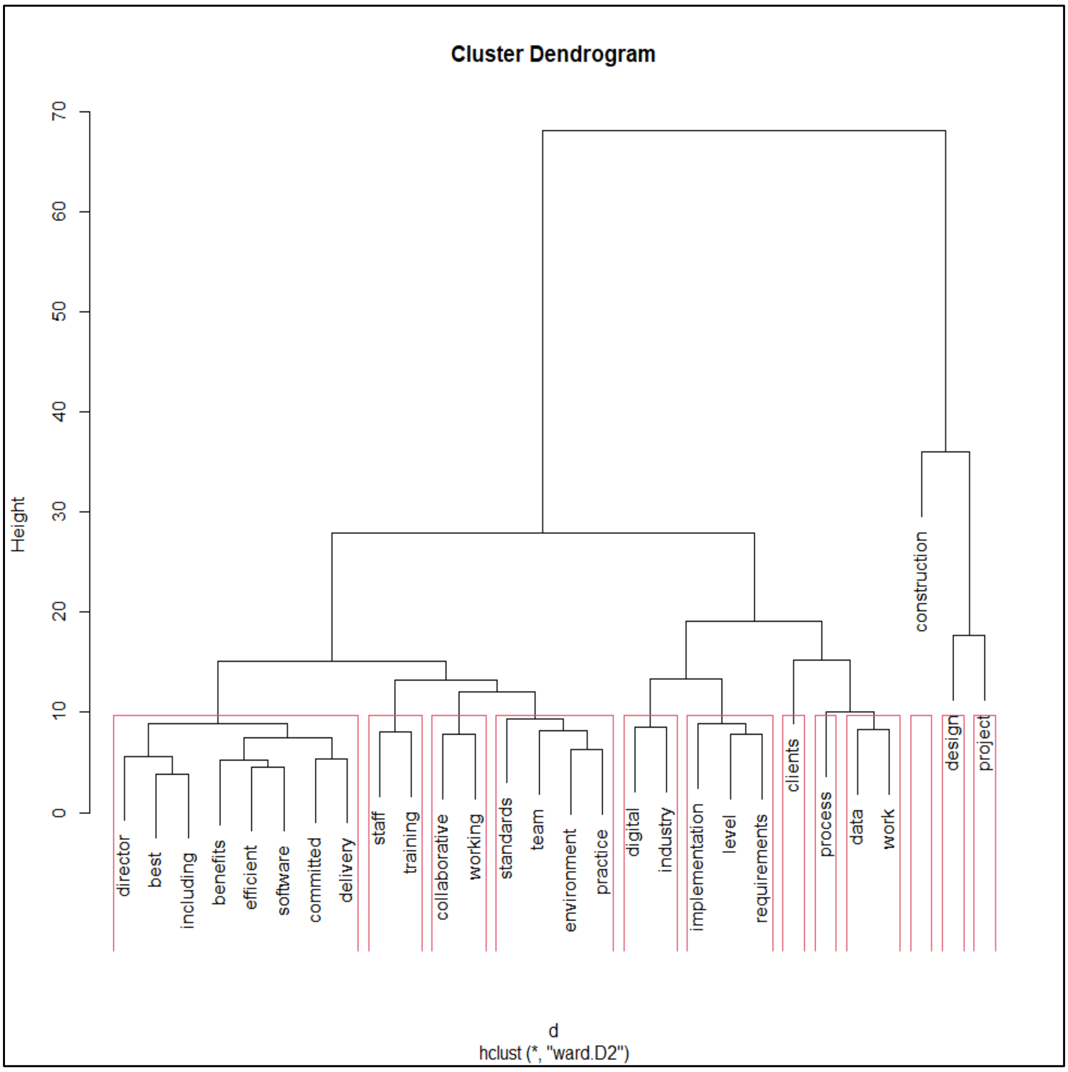


Figure 4.64: Cluster groups dendrogram

The dendrogram presents a visual representation of the clustering of terms obtained from the BIM policy documents. The terms have been grouped into different clusters, indicating their association with each other.

One of the clusters includes the words “implementation”, “level”, and “requirements”, which may suggest that the BIM policies focus on the proper implementation of BIM at various levels, ensuring that the requirements for BIM are met. This cluster highlights the importance of following the established standards and guidelines for implementing BIM.

Another cluster contains the words “collaborative” and “working”, indicating that the BIM policies emphasize the importance of teamwork and collaboration in the BIM process. This is crucial as BIM projects typically involve multiple stakeholders, including designers, contractors, and owners, and their effective collaboration is essential for successful project delivery.

Overall, the clustering of terms reveals valuable insights into the content of the BIM policy documents and what aspects of BIM are considered important by the companies. To further strengthen the findings, the researcher can also conduct a k-means analysis, in the next section, to verify the clustering results.

4.9.2.7 Clustering Documents with the K-Means Algorithm

Table 4.38 provides a clear representation of the five most prominent words in each cluster. This makes it convenient for the researcher to understand the focus of each cluster. By examining the top five words in each cluster, the researcher can gain further insights into the content of the BIM policy documents and the themes that are prevalent within them. The table summarizes the top five words in each of the 12 clusters generated from the clustering analysis of the BIM policy documents. The clusters represent different aspects and themes in the BIM policies.

Table 4.38: The top five words in every cluster

Cluster 1	Clients	Data	Process	Construction	Project
Cluster 2	Construction	Digital	Industry	Level	Collaborative
Cluster 3	Staff	Training	Design	Standards	Clients
Cluster 4	Industry	Digital	Benefits	Delivery	Project
Cluster 5	Project	Work	Construction	Digital	Team
Cluster 6	Project	Clients	Construction	Working	Committed
Cluster 7	Construction	Design	Project	Industry	Process
Cluster 8	Construction	Collaborative	Work	Data	Practice
Cluster 9	Design	Construction	Project	Environment	Data
Cluster 10	Construction	Project	Working	Design	Environment
Cluster 11	Design	Project	Collaborative	Standards	Data
Cluster 12	Project	Implementation	Staff	Process	Collaborative

Cluster 1 emphasizes the role of clients, data, and processes in the construction and project management process. The words “clients”, “data”, and “process” suggest that the BIM policies focus on the data management and process improvement aspect of the construction project.

Cluster 2 focuses on construction, digitalization, and the level of collaboration in the industry. The words “construction”, “digital”, and “collaborative” indicate that the policies aim to integrate digital technologies into construction projects and promote collaboration within the industry.

Cluster 3 highlights the importance of staff training and design standards for clients. The words “staff training”, “design”, and “standards” suggest that the BIM policies prioritize the training of the staff and ensuring that design standards are met to meet the requirements of clients.

Cluster 4 covers the digitalization of the industry and the benefits of project delivery. The words “industry”, “digital”, “benefits”, and “delivery” indicate that the policies focus on the digital transformation of the construction industry and its positive impact on project delivery.

Cluster 5 focuses on the project, team working, and digital construction. The words “project”, “work”, “construction”, and “digital” suggest that the policies aim to promote digital construction practices and improve project outcomes through better team collaboration.

Cluster 6 highlights the commitment of clients and construction projects. The words “project”, “clients”, “construction”, “working”, and “committed” suggest that the policies aim to foster a collaborative relationship between clients and construction teams to achieve project goals.

Cluster 7 covers the relationship between construction, design, and processes. The words “construction”, “design”, “project”, “industry”, and “process” indicate that the policies aim to integrate design and construction processes to achieve efficient and effective outcomes.

Cluster 8 focuses on construction and the practice of collaboration. The words “construction”, “collaborative”, “work”, and “data” suggest that the policies prioritize the collaboration between stakeholders in construction projects to improve outcomes.

Cluster 9 covers the relationship between design, construction, and the environment. The words “design”, “construction”, “project”, “environment”, and “data” suggest that the policies aim to integrate environmental considerations into construction projects and promote sustainable design practices.

Cluster 10 highlights the relationship between construction, project working, and design. The words “construction”, “project”, “working”, “design”, and “environment” indicate that the policies aim to integrate design and construction practices to achieve efficient and effective outcomes in construction projects.

Cluster 11 covers the relationship between design, project, and data. The words “design”, “project”, “collaborative”, “standards”, and “data” suggest that the policies prioritize collaboration and data management to achieve better design outcomes in construction.

The clustering results help to highlight the main areas of focus in the BIM policies, with the first cluster focusing on clients, data, and processes in construction projects, while the second cluster focuses on the benefits and delivery of digital technologies in the construction industry. These results can help to enhance performance in BIM pervasive major project delivery by providing a clear understanding of the areas of focus for BIM policies. This can help to identify areas for improvement and ensure that BIM policies align with the goals and objectives of major construction projects. By focusing on key areas such as data management, process improvement, and digital support, companies can work to improve the efficiency and effectiveness of their BIM implementation and project delivery. The results also highlight the importance of staff training and standards, which can help to ensure that all members of the project team are equipped with the knowledge and skills required to support successful BIM implementation and project delivery.

4.9.3 Lessons Learned from Experiment 5

The analysis of BIM policy documents yielded valuable insights into their coverage and focus areas. The most frequently used terms, including “project”, “construction”, “design”, “clients”, and “process”, indicate that BIM policies aim to encompass all phases of a project and emphasize collaboration among stakeholders. The presence of the term “government” suggests a regulatory aspect to these policies, while words like "engineering", "standards", and "industry" indicate a commitment to adhering to international norms and practices. The word cloud and topic modelling techniques further revealed that key topics within these policies revolve around construction, design, government involvement, and infrastructure. By employing clustering analysis, we were able to group terms related to implementation strategies, collaboration practices, staff training, digitalization efforts, and various aspects of project management. This analysis ultimately provided a comprehensive understanding of the content and themes present in BIM policy documents.

The application of the k-means algorithm to the data resulted in clusters that emphasized different aspects of BIM policy. These clusters highlighted the importance of clients' involvement, data management practices, streamlined processes, digitalization initiatives, adherence to design standards, efficient project delivery methods, collaborative teamwork, environmental considerations, and the promotion of sustainable practices. These findings shed light on the key areas of focus within BIM policies and demonstrate the multifaceted nature of implementing BIM in the construction industry. By considering these insights, policymakers and industry professionals can better understand the core principles and priorities that underpin effective BIM implementation and tailor their strategies accordingly to maximize the benefits of this innovative technology.

4.10 Experiment 6: TS 6 National Policy Papers TM Process

4.10.1 Introduction

The search process for National Policy Papers yielded two Government Construction Strategies (GCS), specifically GCS 2011-2015 and GCS 2016-2020. These strategies, as explained previously, hold significant value in providing insights into the government's approach towards construction and infrastructure development over a specific time period. The retrieval of these GCS documents is justified because they serve as authoritative sources of information regarding the policies, priorities, and objectives set by the government in the realm of construction. By analysing these strategies, policymakers, researchers, and industry professionals can gain a comprehensive understanding of the government's vision for the construction sector, identify trends and shifts in policy focus, and assess the impact of these strategies on the industry's growth and development. Furthermore, studying these GCS documents can inform future policy formulation, guide decision-making processes, and facilitate effective collaboration between the government and stakeholders involved in construction and infrastructure projects. Therefore, the retrieval of the GCS 2011-2015 and GCS 2016-2020 is justified due to their invaluable contribution to shaping the construction landscape and informing policy discourse at a national level.

4.10.2 Pre-processing of the Raw Textual Data

The researcher acquired two GCS documents. In accordance with the methodology, as it was in previous sections, the text was subjected to cleaning processes, followed by various text analysis tasks, with the exclusion of specific terms such as "policy", "building", "information", "modelling", "management", and "BIM" from the corpus. The aim of this process is to ensure consistency in the analysis of the data and to accurately capture the relevant information.

4.10.3 New Knowledge

4.10.3.1 Building A Term-Document Matrix

The researcher obtained a term-document matrix that consisted of 1,796 terms and 2 documents. The matrix was not very sparse (relatively speaking), with only 36% of the entries being zero, as demonstrated in Figure 4.65. In order to gain further insights, the researcher analysed the terms that were used most frequently, as depicted in Figure 4.66.

```
> dtm <- DocumentTermMatrix(docs,)
> dtm
<<DocumentTermMatrix (documents: 2, terms: 1796)>>
Non-/sparse entries: 2301/1291
Sparsity           : 36%
Maximal term length: 99
Weighting          : term frequency (tf)
```

Figure 4.65: Document term matrix (national policy papers)

```
> findFreqTerms(dtm, lowfreq=20)
 [1] "across"        "action"        "actions"
 [4] "agreed"        "approach"      "approaches"
 [7] "appropriate"   "asset"         "benchmarking"
[10] "board"         "cabinet"       "capability"
[13] "carbon"        "central"       "chain"
[16] "client"        "clients"       "construction"
[19] "cost"          "data"          "deliver"
[22] "delivery"      "departmental"  "departments"
[25] "design"        "develop"       "efficiency"
[28] "establish"     "gcs"           "government"
[31] "group"         "identify"      "implementation"
[34] "improve"       "industry"      "infrastructure"
[37] "level"         "mar"           "measures"
[40] "money"         "new"           "objective"
[43] "objectives"    "office"        "ongoing"
[46] "opportunities" "pipeline"      "plan"
[49] "practice"      "process"       "procurement"
[52] "programme"     "progress"      "project"
[55] "public"        "reduction"     "ref"
[58] "review"        "sector"        "skills"
[61] "specific"      "standards"     "strategy"
[64] "summary"       "suppliers"     "supply"
[67] "support"       "targets"       "team"
[70] "theme"         "timescales"    "trial"
[73] "value"         "wholelife"     "work"
[76] "working"
```

Figure 4.66: Top 20 terms

4.10.3.2 Frequent Terms and Associations

A bar plot of the terms that appear with a frequency of at least 20 times in the corpus was achieved by applying certain transformations to remove sparse terms; the plot is shown in Figure 4.67. Additionally, Table 4.39 presents the top 20 most frequently used words in the BIM policy documents.

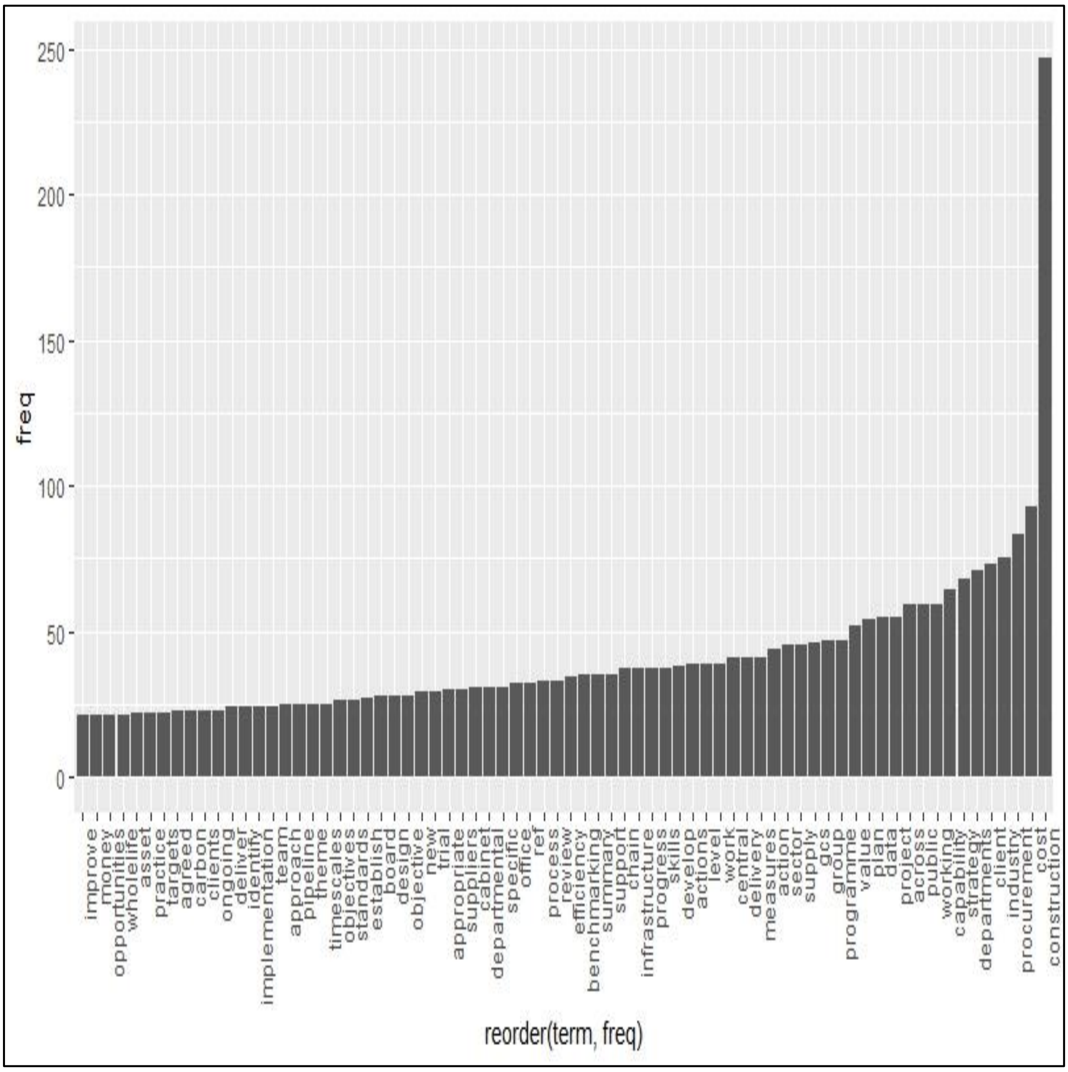


Figure 4.67: Bar plot graph of the frequency of the words

Table 4.39: Frequency of top 20 words

No.	Words	Frequencies
1	Construction	247
2	Cost	93
3	Procurement	83
4	Industry	75
5	Client	73
6	Departments	71
7	Strategy	68
8	Capability	64
9	Working	59
10	Public	59
11	Value	52
12	Programme	47
13	Group	47
14	GCS	46
15	Supply	45
16	Sector	45
17	Action	44
18	Measures	41
19	Delivery	41
20	Central	41

The table shows the 20 most frequently used words in BIM policy documents. The first term, “construction”, appears 247 times, which indicates that the topic of construction is highly emphasized in the BIM policy documents. The second most frequent term is “cost”, which appears 93 times, while the third term, “procurement”, appears 83 times. These three terms reflect the importance of managing costs and procurement processes in the BIM implementation.

The fourth term, “industry”, appears 75 times, and the fifth term, “client”, appears 73 times. These terms show the importance of involving industry stakeholders and clients in the BIM implementation process. The sixth term, “departments”, appears 71 times, highlighting the need for collaboration and coordination among different departments within the government.

The seventh term, “strategy”, appears 68 times, indicating that the development of a comprehensive BIM strategy is important for the successful implementation of BIM in major construction project delivery. The eighth term, “capability”, appears 64 times, emphasizing the need for building capability within the industry to effectively utilize BIM. The ninth term, “working”, appears 59 times, and the tenth term, “public”, appears 59 times, highlighting the importance of collaboration and the involvement of the public sector in the BIM implementation process.

The remaining terms, such as “value”, “program”, “group”, “supply”, “sector”, “action”, “measures”, “delivery”, and “central”, further emphasize the need for effective implementation strategies, coordination, and collaboration among stakeholders, as well as the delivery of value to clients and the public. In conclusion, the results from the analysis of the BIM policy documents highlight the importance of various factors for enhancing performance in BIM pervasive major project delivery, including cost management, procurement processes, stakeholder involvement, comprehensive strategy development, capability building, and effective implementation.

4.10.3.3 Topic Modelling

The top 10 topics discussed in the GCS documents were identified by the researcher and are visualized in a bar graph in Figure 4.68. These topics represent the most frequently discussed subjects in the documents, providing insights into the areas of focus in the national policy papers. The researcher also extracted the top 5 words for each of these topics to further understand the nature of the discussion surrounding each topic. The results of this analysis can be used to understand how the policies support and impact the performance enhancement of BIM in major construction project delivery, providing valuable information for those involved in the construction industry.

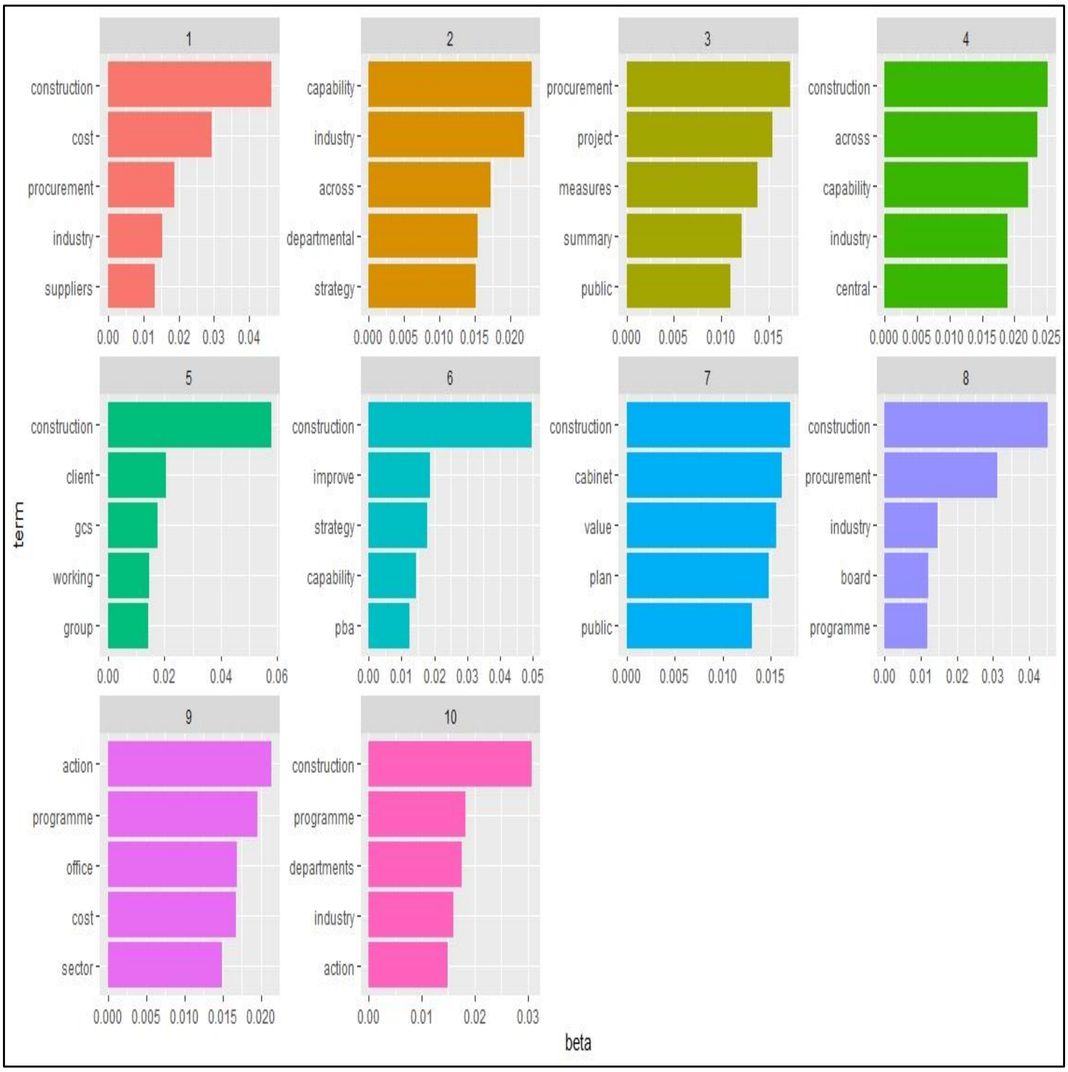


Figure 4.68: Topic modelling output

It is evident from the analysis of the top 10 topics discussed in the GCS documents that the primary focus is on optimizing procurement, cost, and digital capability in the construction industry. Topic 1 highlights the importance of effective procurement and cost-efficient supply management in delivering successful projects. Topic 2 underlines the need to improve digital capabilities across various departments, to ensure better performance and outcomes. Topic 3 emphasizes the significance of performance measurement in public procurement, thereby, ensuring that projects are executed efficiently and effectively. The remaining topics, in general, revolve around cost optimization, procurement strategies, and planning measures, all of which are crucial components of successful project delivery in the BIM pervasive major project delivery domain.

4.10.3.4 Word Cloud

The word cloud, as depicted in Figure 4.69, effectively highlights the key focus areas of the GCS documents, which pertain to performance measurement in public procurement. The most frequently discussed terms include construction, procurement, capability, client, plan, cost, data, and delivery, and further encompasses aspects such as benchmarking, infrastructure, collaboration, innovation, supply chain management, and governance. Additionally, the documents also focus on risks management, efficiency, reduction of carbon emissions, and skill development. These insights provide a comprehensive understanding of the discussions present in the GCS documents and provide valuable information for the improvement of performance in BIM pervasive major project delivery.

Figure 4.69: Word cloud of GCS documents

4.10.4 Knowledge Graphs of TS 3, TS4, TS5, and TS 6

This section synthesises the insights gained from the TM endeavours across TTS 3 to TS 6, revealing the intricate connections between different terms and ideas in the realm of BIM. In the context of the BIMKG and the BIMTAPE Framework, these connections signify a pivotal step towards enhancing performance in major project delivery.

The identified connection between TS5 and TS3 Figure 4.70, encompassing terms such as Work, Design, and Requirements, sheds light on the robust relationship between BIM documentation and the policies guiding BIM implementation within companies. This alignment underscores the convergence of work processes, design principles, and specified requirements, signifying a harmonious integration between technical specifications and organisational BIM policies.

Similarly, the correlation uncovered between TS4 and TS3, focusing on terms like Use, Design, and Requirements, suggests a synchronisation between British Standards (TS4) and fundamental concepts within BIM documentation (TS3). This alignment becomes pivotal in maintaining adherence to established standards and ensuring the seamless integration of BIM processes within the broader industry landscape.

The interconnected terms, such as Team, Project, Data, Construction, Design, and Requirements, in the connection between TS4 and TS5 unveil a substantial overlap between British Standards and BIM policies at companies. This alignment suggests that the standards outlined in TS4 significantly influence the shaping of team dynamics, project management, data handling, construction processes, and design aspects as delineated in BIM policies.

In a parallel vein, the connection between TS4 and TS6, particularly involving terms like Delivery and Construction, signifies that the standards articulated in British Standards (TS4) carry apparent implications for the delivery and construction phases of major projects. This connection underscores the influential role of standards in shaping project delivery processes on a national scale.

The relationship observed between terms such as Working, Industry, and Construction in the connection between TS5 and TS6 underscores a synergistic alignment between organisational BIM policies and broader industry-related policies outlined in national policy papers. This connection emphasises the pivotal role of aligning company practices with industry-wide strategies in the effective implementation of BIM.

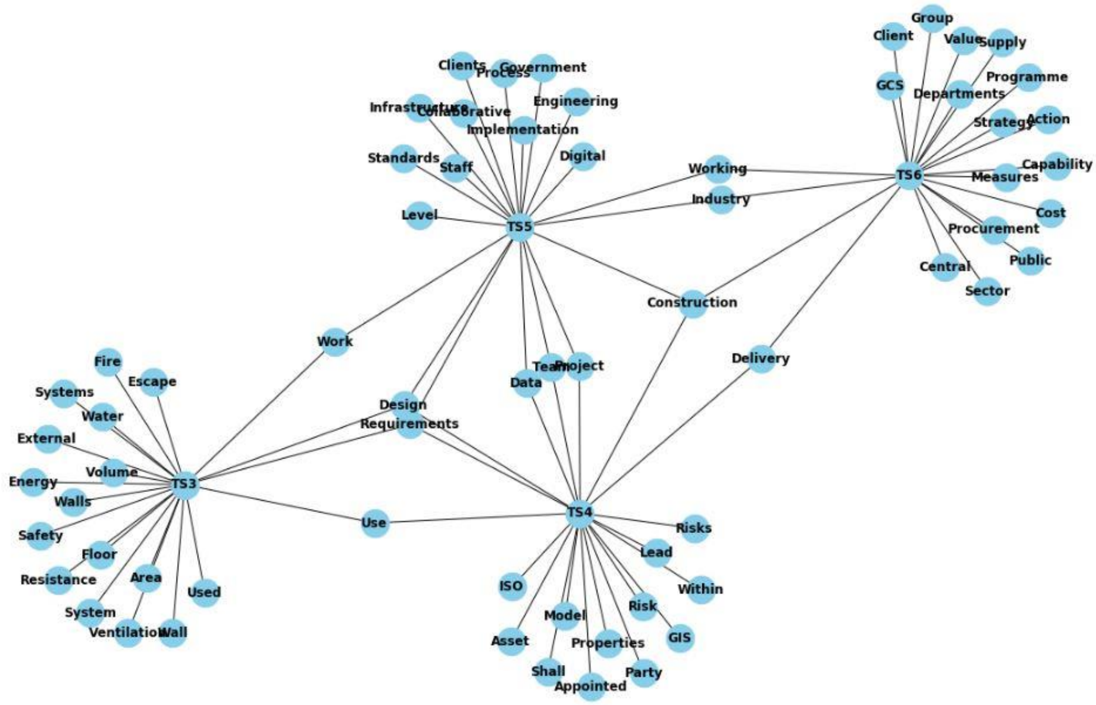


Figure 4.70: Knowledge graph of BIM documentation TS3-TS6

To further elucidate the scope and implications of the identified connections, a comparison table has been crafted (see Table 4.40), showcasing the differences between the prioritized key terms from the Text Analysis process and the comprehensive list of BIM elements outlined in the Periodic Table of BIM across its 10 technical domains (Mordue, 2016). The table serves to delineate clusters of terms in the second column, providing a comprehensive visual aid for understanding the relationship between the two datasets. Each row corresponds to a key term, and the columns represent the BIM domains, showcasing the relationships in a more accessible format.

Table 4.40: Connections between prioritized key terms and the elements in each BIM domain

Periodic Table of BIM Technical Domains	**Clusters of Prioritized Key Terms from Text Analysis**
Strategy	Work, Design, Requirements

Foundations	Use, Design, Requirements
Collaboration	Team, Project, Data, Construction, Design, Requirements
Process	Delivery, Construction
People	Working, Industry, Construction
Technology	Digital
Standards	Standards
Enabling Tools	Collaborative, Process, Working, Digital, Standards, Industry
Resources	Fire, Wall, Water, Floor, Use, Area, Systems, Requirements, Safety, Used, Work, Volume, Ventilation, System, Energy, Escape, Walls, Resistance, External, Design, Shall, ISO, Data, Asset, Delivery, Party, Construction, Risk, Team, Project, Model, Appointed, Risks, Properties, Lead, Within, GIS, Project, Clients, Process, Government, Working, Data, Implementation, Collaborative, Engineering, Industry, Standards, Staff, Level, Infrastructure, Digital, Cost, Procurement, Industry, Client, Departments, Strategy

The Table 4.41 offers a comprehensive reflection on the connections between prioritized key terms and the BIM domain on Strategy, mapping each term to its respective elements within the Periodic Table of BIM. To further enhance understanding, a comparison table can be devised, juxtaposing the complete list of prioritized key terms derived from the Text Analysis process with the comprehensive list of BIM elements organized under the 10 technical domains of the Periodic Table of BIM. This table would serve as a valuable tool for showcasing the scope of the BIMKG in relation to the Periodic Table of BIM, fostering a series of explorations that integrate Text Analysis with Knowledge Graph for BIM-oriented knowledge engineering. This integration aligns with the theory of dependability engineering, providing a robust foundation for dependable BIM implementation. Each BIMKG showcase can contribute to the evolving BIMBOK, aligning with the Academic Interoperability Coalition's standards and practices.

Table 4.41: The connections between prioritized key terms and the BIM domain on Strategy

Priority	**Key Terms from TA Process**	**Reflections to BIM Domain and Elements on Strategy** (Mordue, 2016)
1	Fire	Fire safety and prevention in BIM strategy
2	Wall	Wall design and construction in BIM strategy
3	Water	Water management and utilization in BIM strategy
4	Floor	Floor design and construction in BIM strategy
5	Use	Optimizing space use and functionality in BIM strategy
6	Area	Area measurement and optimization in BIM strategy
7	Systems	Integrated building systems in BIM strategy
8	Requirements	Defining and managing project requirements in BIM strategy
9	Safety	Overall safety considerations in design and construction in BIM strategy
10	Used	Use and repurposing of building elements in BIM strategy
11	Work	Optimizing work processes and efficiency in BIM strategy
12	Volume	Volume calculations and optimization in BIM strategy
13	Ventilation	Ventilation system design and optimization in BIM strategy
14	System	Development and adoption of building systems in BIM strategy
15	Energy	Energy efficiency and consumption optimization in BIM strategy
16	Escape	Designing effective escape routes and systems in BIM strategy
17	Walls	Designing and constructing various types of walls in BIM strategy
18	Resistance	Designing structures with specific resistance properties in BIM strategy
19	External	Optimizing external features and aesthetics in BIM strategy

20	Design	Overall design considerations and coordination in BIM strategy
21	Shall	Compliance with mandatory requirements and regulations in BIM strategy
22	ISO	Adherence to international standards and specifications in BIM strategy
23	Data	Data management, storage, and utilization in BIM strategy
24	Asset	Managing and optimizing building assets in BIM strategy
25	Delivery	Efficient project delivery and completion in BIM strategy
26	Party	Collaboration with involved parties and stakeholders in BIM strategy
27	Construction	Effective and efficient construction processes in BIM strategy
28	Risk	Identifying, assessing, and managing project risks in BIM strategy
29	Requirements	Aligning with and fulfilling project requirements in BIM strategy
30	Design	Comprehensive and coordinated design processes in BIM strategy
31	Team	Fostering collaboration and communication within project teams in BIM strategy
32	Project	Overall project management and execution in BIM strategy
33	Use	Optimizing space use and functionality (repeated for clarity) in BIM strategy
34	Model	Creating, managing, and utilizing digital models in BIM strategy
35	Appointed	Appointing key roles and responsibilities within a project in BIM strategy
36	Risks	Identifying, assessing, and mitigating risks in project execution in BIM strategy
37	Properties	Managing and optimizing building properties and characteristics in BIM strategy
38	Lead	Providing leadership and direction within project teams in BIM strategy

39	Within	Considerations within specific project contexts in BIM strategy
40	GIS	Integrating Geographic Information Systems (GIS) data in BIM strategy
41	Project	Overall project management and execution (repeated for clarity) in BIM strategy
42	Construction	Effective and efficient construction processes (repeated for clarity) in BIM strategy
43	Design	Comprehensive and coordinated design processes (repeated for clarity) in BIM strategy
44	Clients	Understanding and meeting client needs and expectations in BIM strategy
45	Process	Optimizing and streamlining project processes in BIM strategy
46	Government	Compliance with governmental regulations and standards in BIM strategy
47	Work	Optimizing work processes and efficiency (repeated for clarity) in BIM strategy
48	Data	Data management, storage, and utilization (repeated for clarity) in BIM strategy
49	Implementation	Implementing and integrating BIM processes and technologies in BIM strategy
50	Working	Fostering collaborative and effective working environments (repeated for clarity) in BIM strategy
51	Collaborative	Collaboration and coordination among project stakeholders in BIM strategy
52	Engineering	Applying engineering principles and practices to project design in BIM strategy
53	Industry	Aligning with and contributing to industry standards and practices in BIM strategy
54	Standards	Adherence to BIM standards and industry best practices in BIM strategy
55	Staff	Developing and managing project staff, skills, and expertise in BIM strategy

56	Level	Considering and optimizing building levels and elevations in BIM strategy
57	Infrastructure	Planning and optimizing building infrastructure and systems in BIM strategy
58	Team	Fostering collaboration and communication within project teams (repeated for clarity) in BIM strategy
59	Requirements	Aligning with and fulfilling project requirements (repeated for clarity) in BIM strategy
60	Digital	Leveraging digital technologies and tools for project optimization in BIM strategy
61	Construction	Effective and efficient construction processes (repeated for clarity) in BIM strategy
62	Cost	Managing and optimizing project costs and expenditures in BIM strategy
63	Procurement	Efficient procurement and sourcing of project materials and services in BIM strategy
64	Industry	Aligning with and contributing to industry standards and practices (repeated for clarity) in BIM strategy
65	Client	Understanding and meeting client needs and expectations (repeated for clarity) in BIM strategy
66	Departments	Coordinating and aligning with various project departments and functions in BIM strategy
67	Strategy	Overall project strategy and planning in BIM strategy
68	Capability	Assessing and enhancing organizational capabilities for BIM implementation in BIM strategy
69	Working	Fostering collaborative and effective working environments (repeated for clarity) in BIM strategy
70	Public	Considerations related to public interests and engagement in BIM strategy
71	Value	Maximizing value and benefits in project delivery in BIM strategy
72	Programme	Managing and optimizing project programs and schedules in BIM strategy
73	Group	Collaboration and coordination within project groups and teams in BIM strategy

74	GCS	Adhering to the UK Government Construction Strategy in BIM strategy
75	Supply	Managing and optimizing the supply chain in project delivery in BIM strategy
76	Sector	Considerations specific to the project sector or industry in BIM strategy
77	Action	Taking actionable steps and initiatives in project delivery in BIM strategy
78	Measures	Defining and implementing performance measures in project delivery in BIM strategy
79	Delivery	Efficient project delivery and completion (repeated for clarity) in BIM strategy
80	Central	Centralized coordination and management of project activities in BIM strategy

The technical advantages of BIMKG lie in its dynamic nature, enabling the Periodic Table of BIM to adapt and support real-world BIM practices. This integration facilitates a seamless connection between theoretical knowledge and practical application, enriching the BIMBOK. Moreover, the BIMKG, showcased across experiments or as a holistic entity, plays a pivotal role in making the Periodic Table of BIM a dynamic tool for practitioners. This aligns with the Academic Interoperability Coalition's efforts to establish a comprehensive Body of Knowledge (BOK) for BIM, enhancing interoperability and knowledge exchange in the academic community.

The powerful integration of the BIMTAPE Framework with BIM practice, as summarized by the Periodic Table of BIM, represents a significant step towards aligning industry standards with practical implementation. This synergy ensures that BIM processes, guided by the standards outlined in the Periodic Table, seamlessly adhere to the principles of the BIMTAPE Framework. The collaborative framework presented in the Periodic Table becomes a practical guide for major project delivery, reinforcing the importance of coherence and alignment between different facets of BIM, from standards to company policies and national guidelines. This integration not only enhances performance in major project delivery but also contributes to the evolution of best practices in the BIM ecosystem.

4.10.5 Lessons Learned from Experiment 6

The analysis of the term-document matrix and topic modelling in the context of the GCS documents yielded valuable lessons for enhancing performance in BIM pervasive major project delivery. The examination of the most frequently used terms in the BIM policy documents revealed key areas of emphasis, such as construction, cost, procurement, industry, client, and strategy. These findings emphasize the importance of managing costs, involving industry stakeholders and clients, developing comprehensive strategies, and building capability within the construction sector to effectively implement BIM. Collaboration and coordination among different departments within the government were also highlighted as essential factors. The results underscore the significance of factors such as cost management, procurement processes, stakeholder involvement, strategy development, capability building, and effective implementation in driving performance improvement in BIM-major construction project delivery.

Additionally, the topic modelling analysis identified the top 10 subjects discussed in the GCS documents. These topics encompassed optimizing procurement, cost, and digital capability in the construction industry. The focus on effective procurement, cost-efficient supply management, and improving digital capabilities across departments indicated the importance of these aspects in achieving successful project outcomes. Furthermore, topics related to performance measurement in public procurement, cost optimization, procurement strategies, and planning measures were found to be integral to BIM pervasive major project delivery. These insights provide valuable guidance for policymakers, researchers, and industry professionals involved in the construction sector, highlighting the areas of focus in national policy papers and aiding in the development of strategies for performance enhancement.

In conclusion, the analysis of the term-document matrix, frequent terms, topic modelling, and word cloud in the context of BIM policy documents yielded important lessons for improving performance in BIM pervasive major project delivery. The findings underscore the significance of managing costs, involving stakeholders, developing comprehensive strategies, building capability, and effectively implementing BIM. The focus on optimizing procurement, cost, and digital capability within the construction industry is crucial for successful project outcomes. These insights provide valuable guidance for policymakers and industry professionals, informing policy development and supporting performance enhancement in BIM-major construction project delivery.

In the integration of TM and the BIMKG, insights were synthesized across TS 3 to TS 6, revealing intricate connections between terms and ideas in the realm of BIM. Notably, the alignment between technical specifications and organisational BIM policies (TS5 and TS3) and the synchronisation between British Standards and fundamental BIM concepts (TS4 and TS3) emphasise the harmonious integration crucial for major project delivery. The connections between TS4 and TS5, TS4 and TS6, and TS5 and TS6 underscore the significant influence of British Standards on team dynamics, project management, data handling, construction processes, and industry-wide strategies. A comparison table further elucidates the relationships between prioritised key terms and the elements in each BIM domain, highlighting the comprehensive scope of the BIMKG and its integration with the Periodic Table of BIM. This integration, aligned with dependability engineering theory, enhances BIM practices' real-world adaptability, contributing to the evolving Body of Knowledge for BIM and fostering interoperability in the academic community. The synthesis concludes by emphasising the powerful integration of the BIMTAPE Framework with BIM practice, as showcased in the Periodic Table of BIM, representing a significant stride towards aligning industry standards with practical implementation and promoting best practices in the BIM pervasive major project delivery.

4.11 Validation of BIM TM Approach

4.11.1 Introduction

The BIMTAPE Framework model validation is a crucial step in assessing the effectiveness and performance of the devised TM model in achieving the research objectives pertaining to TM and analysis. In this validation process, SNA plays a vital role in evaluating the model’s ability to categorize texts, extract meaningful information, and provide valuable insights. By applying SNA techniques, the validation aims to examine the relationships, connections, and interactions between words and topics/classes within the BIM model. Through a comparative analysis and manual verification, the validation process determines the accuracy of the TM model in mapping words to specific topics/classes. Ultimately, the validation using SNA contributes to the development of an effective TM approach for enhancing the performance of major construction project delivery in the context of BIM. The iterative nature of the process ensures continuous improvement and refinement of the model's results.

4.11.2 Social Network Analysis

SNA is a research methodology used for exploring the structure and relationships within a group of individuals or organizations (Glegg et al., 2019). SNA has been recognized as a useful tool for validating the outcomes of TM, which is a process of extracting and analysing text data to uncover patterns and trends. This method of analysis examines both the structure of relationships between the terms, as well as the natural connections that occur within them.

In SNA, nodes are visually represented as objects, while links symbolize the relationships, associations, or modes of interaction between the nodes (Pryke, 2012). SNA provides a quantitative and visual representation of the interactions and relationships, making it easier to comprehend and understand the findings. SNA has been proven to be a powerful tool for studying the interactions and relationships between objects and can be applied in various domains (Glegg et al., 2019; Scott et al., 2005). SNA is utilized as a validation technique for TM results because it enables researchers to examine the connections and interactions between the terms in a visual and quantitative manner. This helps to ensure that the results of TM are accurate and representative of the underlying structure of relationships between the terms.

By analysing the connections between the nodes in the network, SNA can highlight any discrepancies or outliers in the data and provide a more comprehensive understanding of the relationships between the terms. Additionally, SNA can be useful in validating TM results by providing a visual representation of the relationships between the terms, allowing researchers to easily identify patterns and trends in the data. Ultimately, the combination of TM and SNA can lead to a deeper understanding of the underlying structure of relationships within a group of people or organizations, helping researchers to draw insights and make informed decisions based on their findings.

The utilization of SNA as a validation technique for TM results can provide a comprehensive understanding of the relationships between different entities, such as the word frequencies in each experiment or the clustering of words that reveal trends. For instance, if the TM process has identified a group of words that are commonly co-occurring, SNA can be utilized to create a network diagram that exhibits the connections between these words. The application of SNA in validating TM results can assist in the identification of patterns and relationships within the data, and also provide a visual representation of the data that can be easily comprehended and analysed, thereby improving the validity of the TM results.

As a result, the SNA theory is utilized as an evaluation tool to analyse the results of experiments in a quantitative and visual manner by identifying the co-occurrence of terms in documents. In this study, the nodes represent the various components, and the links represent their associations. This forms the six dimensions that were previously discussed in the experiments, except for the BIM TS 2 experiment, which was conducted in a real-time scenario. This research not only expands the knowledge base in two rapidly growing areas of research and practice, but also highlights the potential to inform further research and development with regards to managing construction projects throughout their lifecycle at the tactical and operational level. With the growing challenges posed by the adoption of BIM, this new socio-technical subsystem must be well managed in construction project delivery.

4.11.3 Building Relationship Graphs

The first step in this process involves the importation of a term-document matrix into the R software environment (as depicted in Figure 4.71). The term-document matrix contains the 20 most frequent terms that have been extracted from each experiment (as previously discussed). The data is then transformed into a term-term adjacency matrix, serving as the basis for the construction of a graph. The graph is then visualized to demonstrate the relationships between the frequent terms, with the use of color, font size, and vertex and edge transparency adjustments to improve readability. Through the systematic analysis of these terms, the present study aims to offer knowledge-based findings that can inform the development of a roadmap for future research and guide the efforts of researchers and professionals in enhancing performance in BIM pervasive major project delivery through the application of TM techniques.

```
> #Social network analysi
> #load TermDocumentMatrix
> library(readr)
> termDocMatrix <- read.csv("~/Most 20 FreqTerms.rdata")
> # inspect part of the matrix
> termDocMatrix[1:5, 5:10]
  Floor  Use Area Systems Requirements Safety
1  1331 1200 1065    1877          955    937
2     0  290    0       0          308      0
3     0    0    0       0           38      0
4     0    0    0    2956            0   1240
5     0    0    0       0            0      0
> termDocMatrix <- as.matrix(termDocMatrix)
> # change it to a Boolean matrix
> termDocMatrix[termDocMatrix>=1] <- 1
> # transform into a term-term adjacency matrix
> termMatrix <- termDocMatrix
> # inspect terms numbered 5 to 10
> termMatrix[1:5,5:10]
     Floor  Use    Area   Systems Requirements Safety
[1,] "1"    "1"    "1"    "1"     "1"          " 937"
[2,] "   0" " 290" "   0" "   0" "1"          "   0"
[3,] "   0" "   0" "   0" "   0" "  38"       "   0"
[4,] "   0" "   0" "   0" "1"    "   0"       "1"
[5,] "   0" "   0" "   0" "   0" "   0"       "   0"
> dim(termMatrix)
[1] 73 73
```

Figure 4.71: Loading term document matrix into Rstudio

An analysis of the words from each experiment was performed and it was observed that the frequency of words aligns with the results obtained from the previously described matrix. For instance, the term “safety” in the approved documents experiment was recorded to have been mentioned 937 times. This confirms the validity and completeness of the loaded data. The term-document matrix serves as the result of the product of two matrices, with the transposition of the matrix being applied. Subsequently, a term-term adjacency matrix was created, where the rows and columns correspond to terms, and each entry represents the number of co-occurrences between the two terms. The graph can be constructed using the ‘graph.adjacency()’ function from the ‘igraph’ package (as depicted in Figure 4.72). The code snippet was developed by the author, building on examples from existing resources, and utilizes the ‘igraph’ library in R (Csardi and Nepusz, 2006) to construct and simplify a graph based on a term matrix, enabling the visualization of term relationships. Following this step, the researcher plots the network using the 'layout.fruchterman.reingold' function (as illustrated in Figure 4.73).

```
> library(igraph)
> # build a graph from the above matrix
> g <- graph.adjacency(termMatrix, weighted=T, mode = "undirected")
> # remove loops
> g <- simplify(g)
> # set labels and degrees of vertices
> V(g)$label <- V(g)$name
> V(g)$degree <- degree(g)
```

```
> # set seed to make the layout reproducible
> set.seed(3952)
> layout1 <- layout.fruchterman.reingold(g)
> plot(g, layout=layout1)
```

Figure 4.72: A code for building a graph in R (developed by the author)

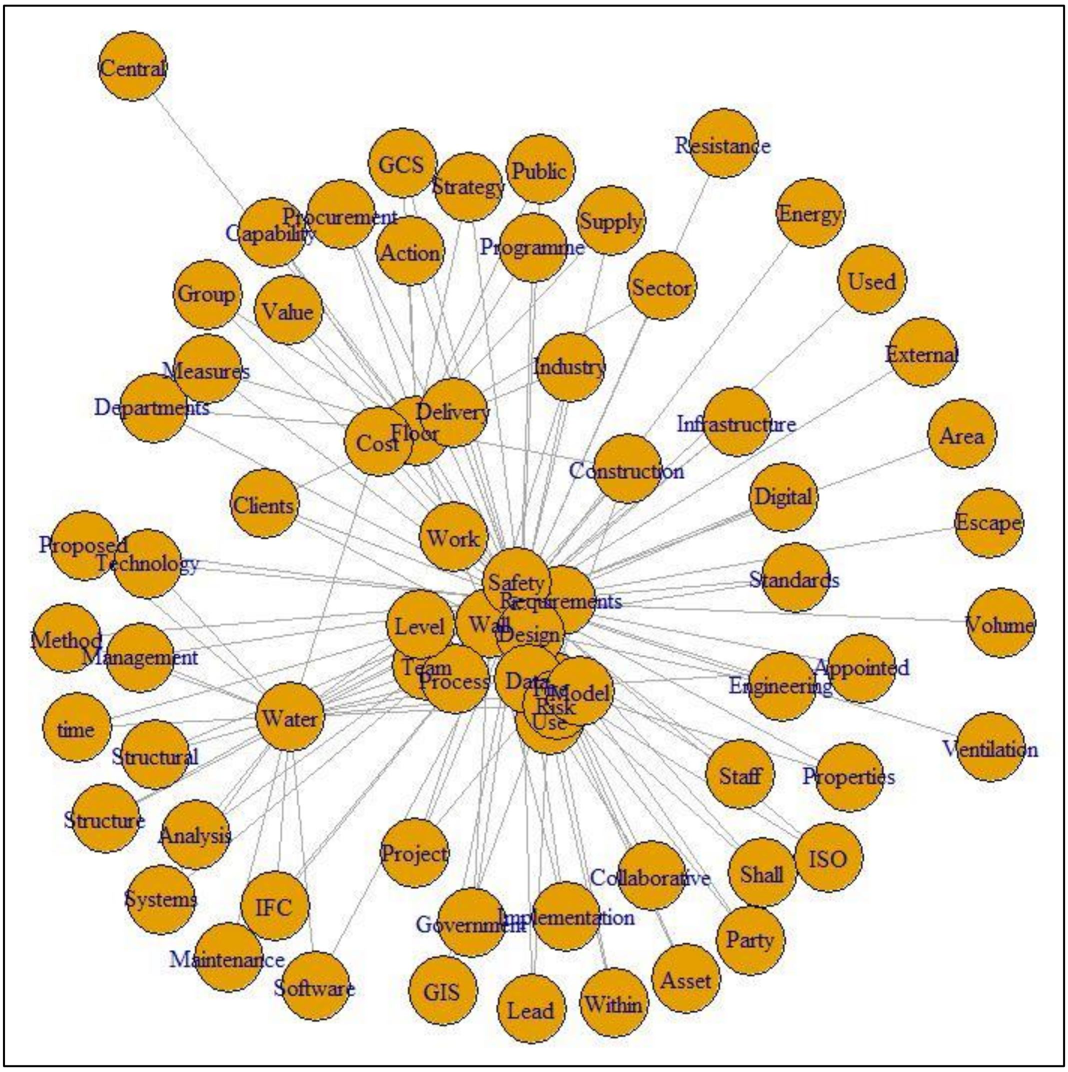


Figure 4.73: Network layout

In the aforementioned code, the layout is stored as ‘layout1’, to ensure that the graph can be plotted in the same layout at a later stage. An alternative layout can be generated through the execution of the first line of code below (Figure 4.74). The second line of code generates an interactive plot, providing the researcher with the ability to manually rearrange the layout (as shown in Figure 4.75).

```
> plot(g, layout=layout.kamada.kawai)
> tkplot(g, layout=layout.kamada.kawai)
```

Figure 4.74: Code for generating layout

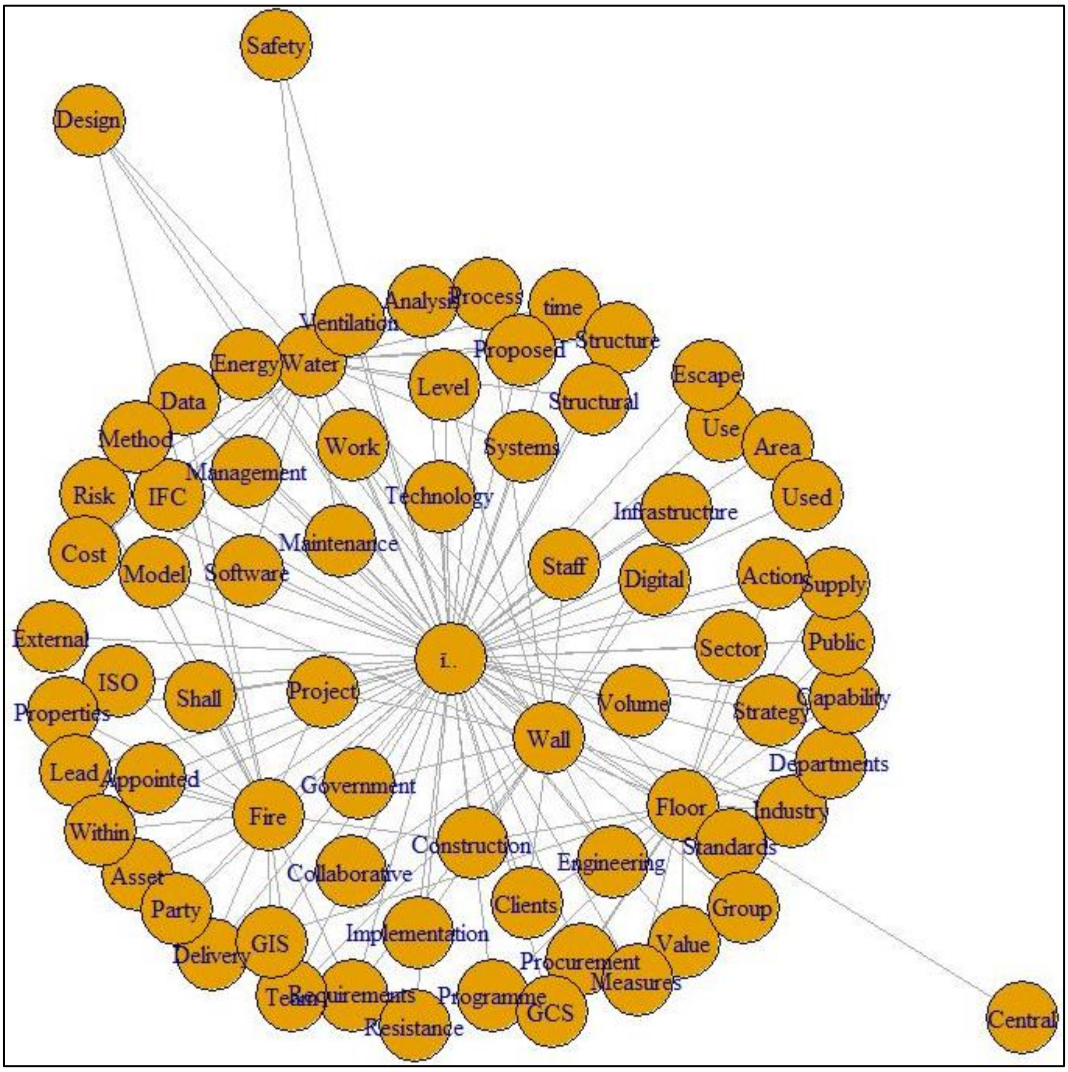


Figure 4.75: New layout

Subsequently, the researcher adjusts the label size of vertices based on their degrees to emphasize important terms. In a similar manner, the researcher also sets the width and transparency of edges based on their weights. This is particularly useful in situations where graphs have a large number of vertices and edges. In the code below, the vertices and edges are accessed using the ‘V()’ and ‘E()’ functions (as depicted in Figure 4.76). The ‘rgb(red, green, blue, alpha)’ function defines a colour with a specified alpha transparency (as illustrated in Figure 4.77).

```
> V(g)$label.cex <- 8.8 * V(g)$degree / max(V(g)$degree)+ .8
> V(g)$label.color <- rgb(0, 0, .5, .8)
> V(g)$frame.color <- NA
> egam <- (log(E(g)$weight)+.8) / max(log(E(g)$weight)+.8)
> E(g)$color <- rgb(.4, .4, 0, egam)
> E(g)$width <- egam
> # plot the graph in layout1
> plot(g, layout=layout1)
```

Figure 4.76: A code for vertices and edges

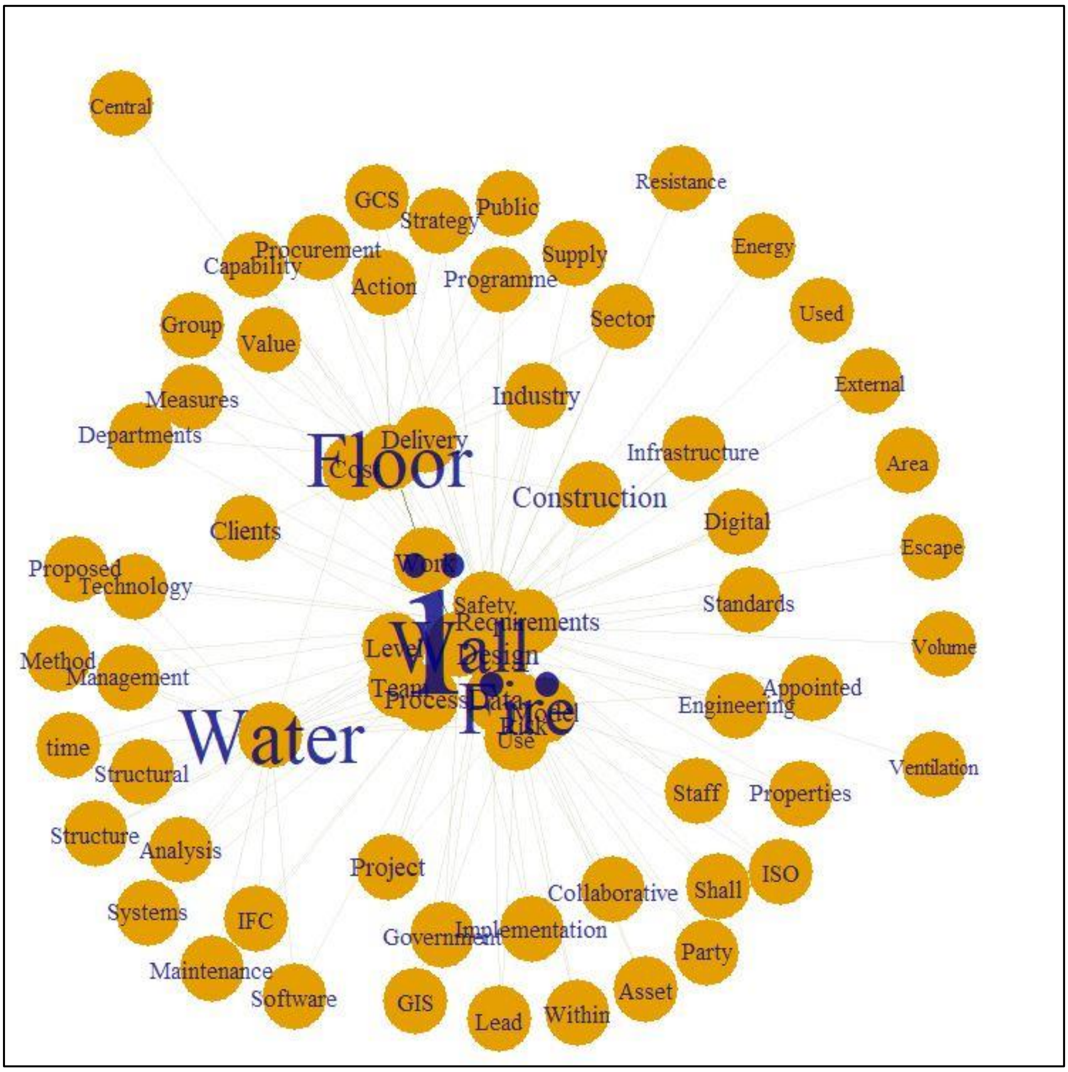


Figure 4.77: New created network

4.11.4 Network of Top 20 Frequent Words

In the process of analysing a dataset, the researcher has employed a graphical representation approach to depict the relationship between the most frequent words in the dataset. To construct the graph, the researcher utilized the frequency of terms that the words shared, as a basis for determining the connections between the words. The result is a graph of the 20 most frequent words, as depicted in Figure 4.78. It can be seen that the graph is highly populated due to the presence of common words among the frequent words, leading to numerous connections between words. To simplify the graph and improve its interpretability, the researcher conducted an analysis of the distribution of vertex degrees, which is a measure of the number of connections a vertex has in a graph. The results of this analysis are presented in Figure 4.79.

The analysis revealed that there are approximately 45 vertices with a degree of two, which implies that these vertices have only two connections in the graph. This information is valuable in understanding the relationships between the words in the dataset and can provide insights into the underlying structure of the data.

```
> M <- termDocMatrix
> BIMmatrix <- M
> library(igraph)
> g <- graph.adjacency(BIMmatrix, weighted=T,
+                      mode="undirected")
> V(g)$degree <- degree(g)
> g <- simplify(g)
> V(g)$label <- V(g)$name
> V(g)$label.cex <- 1
> V(g)$label.color <- rgb(.4, 0, 0, .7)
> V(g)$size <- 2
> V(g)$frame.color <- NA
> barplot(table(V(g)$degree))
```

Figure 4.78: A graph of top 20 frequent words

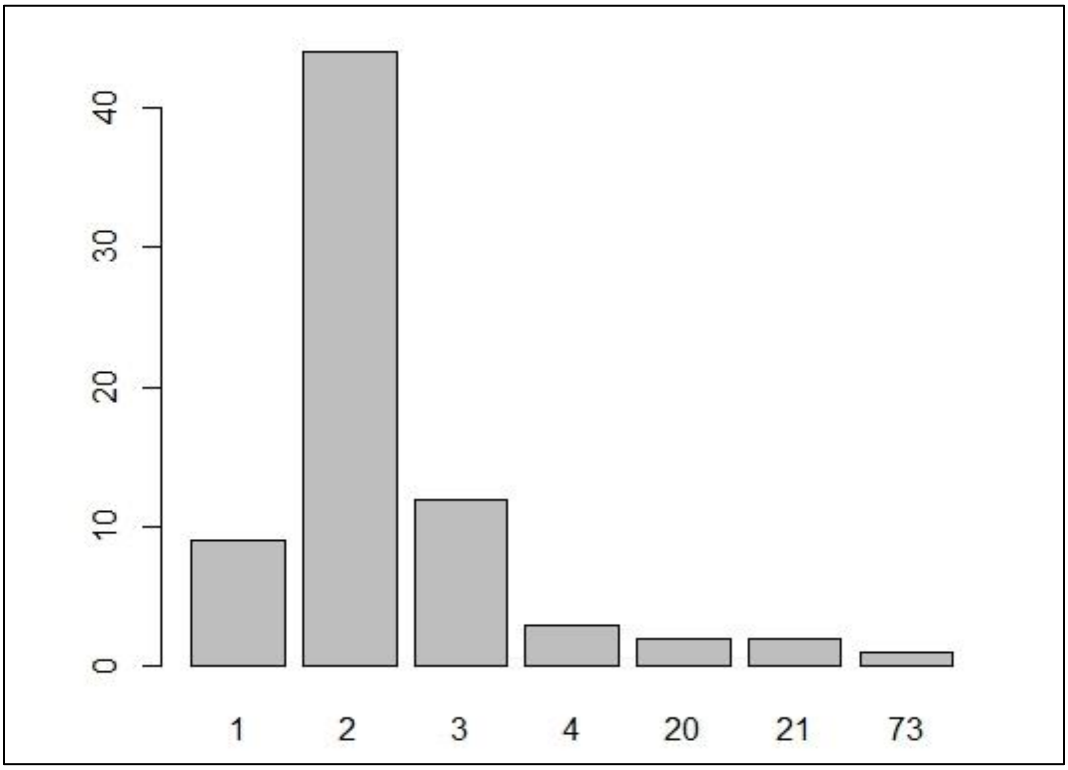


Figure 4.79: The distribution of degree of vertices

By utilizing the code provided, the researcher assigned colours to the vertices based on their degree and labelled the isolated vertices with both the document ID and the first ten characters of each word (as depicted in Figure 4.80). To maintain clarity and prevent label overcrowding, the researcher designated the labels of other vertices as solely the document ID. Furthermore, the colour and width of the edges were adjusted according to their weight, resulting in a graph with improved visual representation (as shown in Figure 4.81).

```
> idx <- V(g)$degree == 0
> V(g)$label.color[idx] <- rgb(0, 0, .3, .7)
> df <- do.call('rbind', lapply(termDocMatrix, as.data.frame))
> V(g)$label[idx] <- paste (V(g)$name[idx], substr(df$text[idx], 1, 10), sep=":")
> egam <- (log (E(g)$weight)+.2) / max(log (E(g)$weight)+.2)
> E(g)$color <- rgb(.5, .5, 0, egam)
> E(g)$width <- egam
> set.seed (3152)
> layout2 <- layout.fruchterman.reingold(g)
> plot(g, layout=layout2)
```

Figure 4.80: Setting vertex colours based on degree

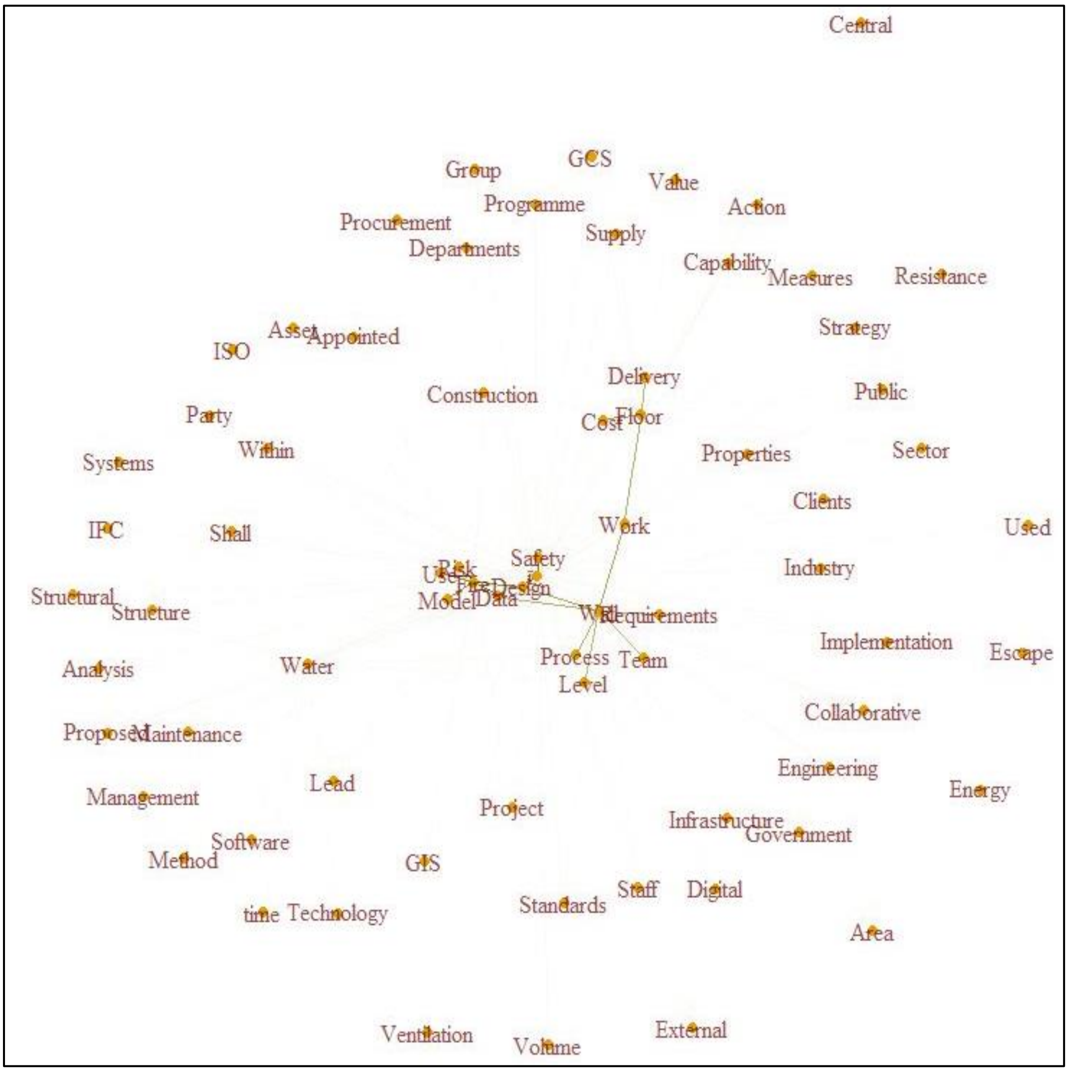


Figure 4.81: Setting the colour and width of edges

As evident in Figure 4.81, some vertices (in a crescent) are isolated and not connected to any others. In order to simplify the graph and remove these isolated vertices, the researcher employed the 'delete.vertices()' function to remove them (see Figure 4.82) and then replotted the graph (as shown in Figure 4.83).

```
> g2 <- delete.vertices(g,V(g)[degree(g) ==0])
> plot(g2, layout=layout.fruchterman.reingold)
```

Figure 4.82: Code to remove isolated vertices

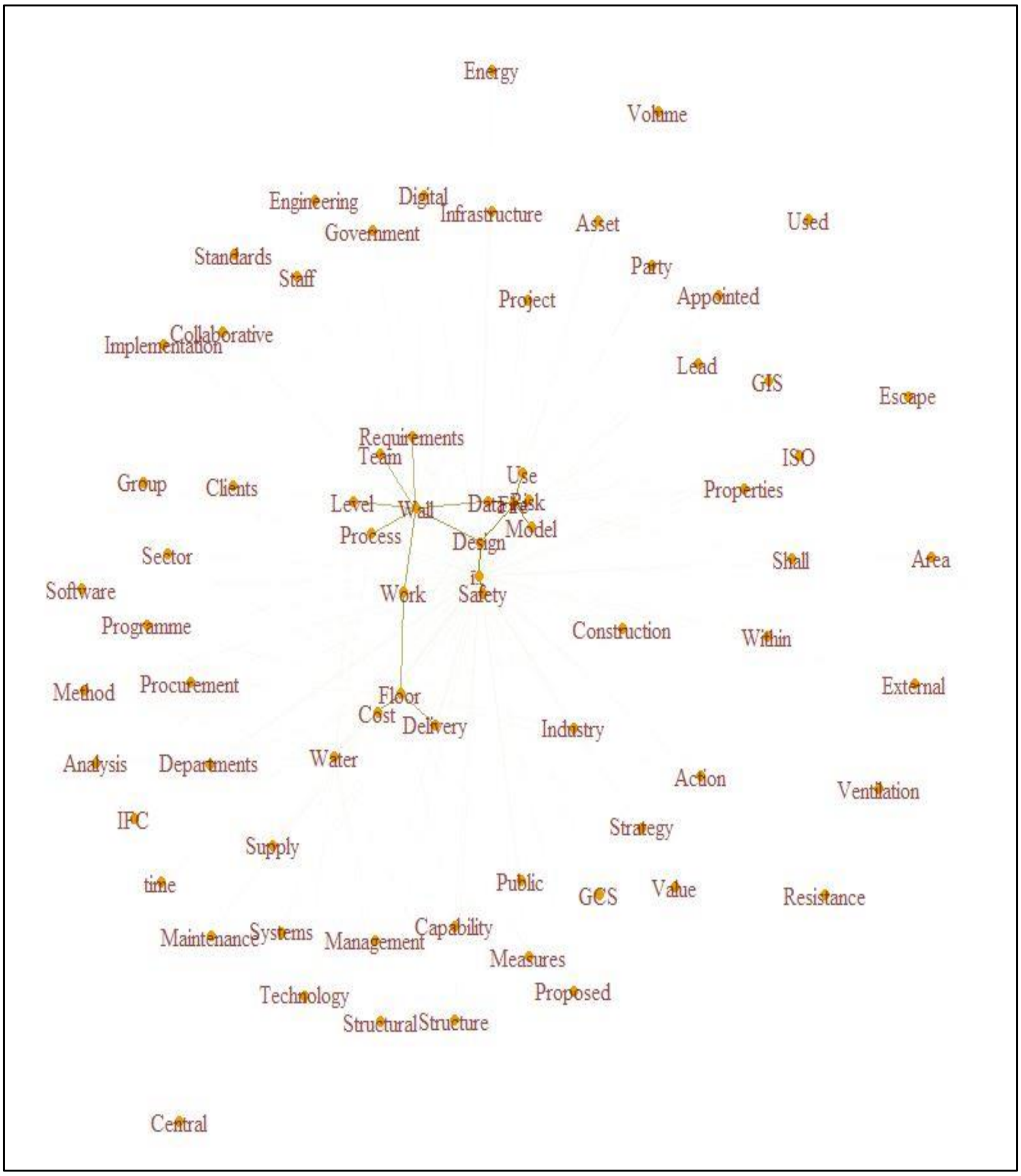


Figure 4.83: Network after removing isolated vertices

In a similar manner, the researcher has also removed edges with low degrees to simplify the graph. Using the ‘delete.edges()’ function, the researcher removed edges with a weight of one (as illustrated in Figure 4.84). Following the removal of these edges, some vertices became isolated, and these were also removed (as depicted in Figure 4.85).

```
> g3 <- delete.edges(g, E(g)[E(g)$weight <= 1])
> g3 <- delete.vertices (g3, V(g3)[degree(g3) == 0])
> plot (g3, layout=layout.fruchterman.reingold)
```

Figure 4.84: A code for removing edges

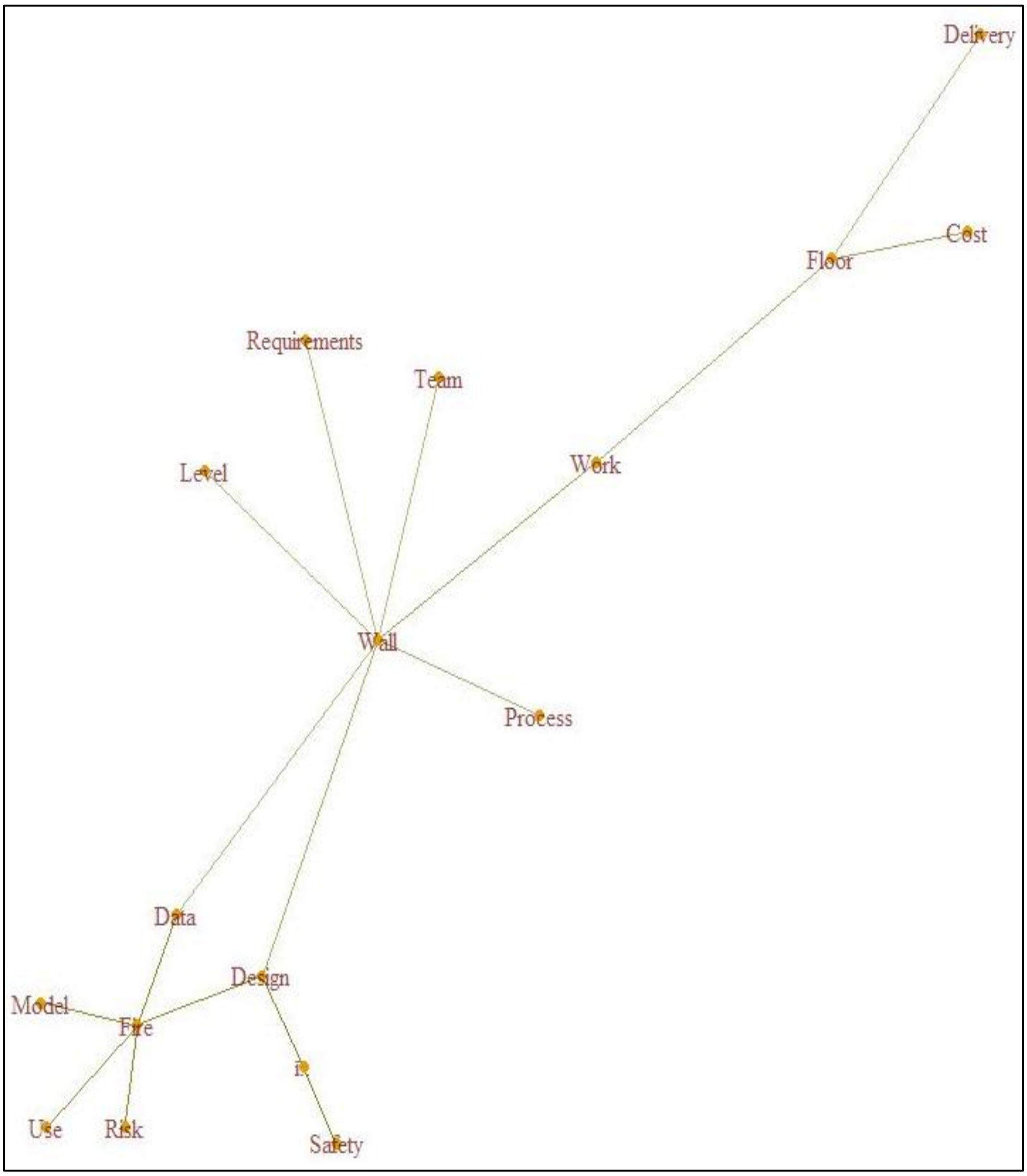


Figure 4.85: Graph with edges removed

The visualization presented in Figure 4.85 is an outcome of the simplification process carried out by the researcher on the graph of the most frequent words. This process revealed the presence of distinct clusters of words that are related to each other in some way. Each of these clusters, or cliques, contains words that seem to share a common theme or topic.

The first clique, located in the lower left portion of the figure, encompasses words such as “fire”, “data”, “design”, “safety”, “risk”, “use”, and “model”. These words are likely related to safety and risk management and could indicate that these topics are frequently discussed in the documents being analysed.

The second clique, located in the central region of the figure, consists of words like “wall”, “process”, “team”, “requirements”, and “level”. These words are associated with process management, teamwork, and the fulfilment of requirements. This clique suggests that these topics are also discussed frequently in the documents.

The third clique, located in the top right corner of the figure, comprises of three interconnected words, which are “floor”, “cost”, and “delivery”. These words could be related to construction or engineering and might indicate that these topics are discussed in the analysed documents. The word “wall” acts as a connector between these three cliques, implying that the word “wall” may be related to both safety and construction-related topics.

This analysis highlights the usefulness of visualizing the relationships between words in a document or set of documents. It allows the researchers to quickly identify common themes and topics discussed in the analysed data, which can be useful in various applications such as text analysis and information retrieval.

4.11.5 Two-Mode Network

In this section, the researcher will construct a two-mode network which consists of two types of vertices: “documents”, represented by “experiments”, and “terms”, represented by the 20 most frequent words in each experiment. To begin, the researcher directly generates a graph, “g”, from the term-document matrix. Subsequently, the software NetDraw (Borgatti et al., 2002) is utilized to perform graphical representation and interpretation of the communication networks. Concepts from SNA, including path length, actor centrality, network density, and tie strength, are employed for comparative analysis of the projects, as depicted in Figure 4.86.

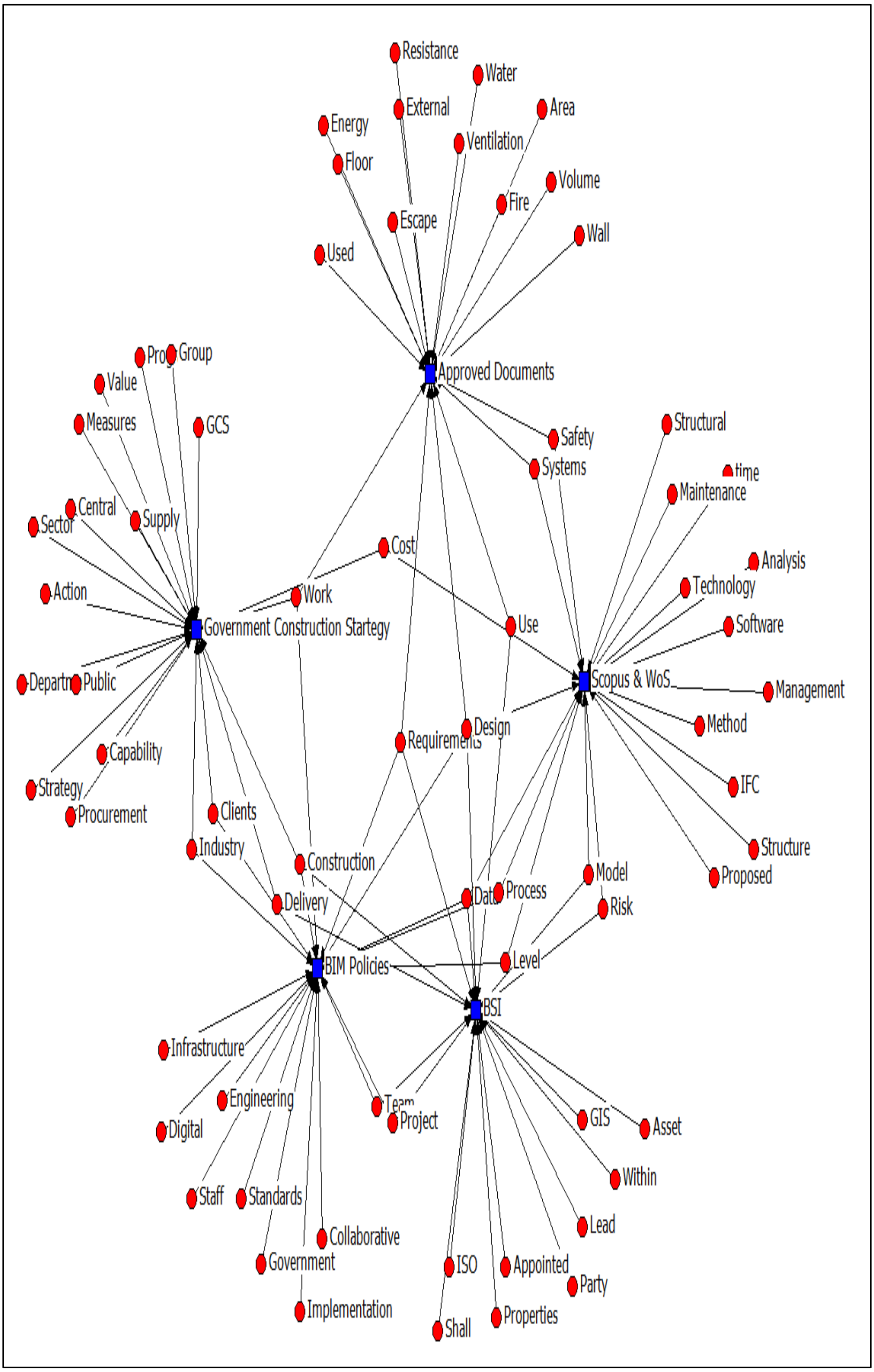


Figure 4.86: Two-mode network relationships

Figure 4.86 depicts a two-mode network of relationships between five experiments and the 20 most frequent words in each experiment. The blue dots represent the experiment names, and the arrows represent the relationships and the degree of interaction between the words in the experiments. The network encompasses all words that emerged from the experiments in the study, and the relationships between them. The ties between the words are based on the strength of their relationship, as indicated by a tie strength score. It can be observed that all dots are connected, suggesting that there exists some sort of connection among the studied documents.

In the analysis of the two-mode network, the researcher employed the concept of k-core, which is a subgraph in which all vertices have a minimum degree of k. The k-core is a measure of the centrality of a vertex in a network, indicating the level of connectivity and influence it has within the graph. To calculate the k-core, the process described by Batagelj and Zaversnik (2003) was employed, whereby an algorithm was applied to the graph that iteratively removes vertices with a degree less than k until only vertices with degrees equal to or greater than k remains. The subgraph that results from this process is considered the k-core of the graph.

Batagelj and Zaversnik (2003) observed that the calculation of k-cores is a useful tool in SNA for identifying central or influential nodes within a network. The k-core of a network is a subgraph in which all vertices have a minimum degree of k. The calculation of k-cores involves iteratively removing vertices with a degree less than k until no such vertices remain, resulting in a subgraph that represents the k-core of the network. These k-cores can be used to identify the most tightly connected group of vertices within a network, as well as to understand the overall structure of the network and to identify key players. The application of k-cores in a network analysis in this case can be seen in Figure 4.87.

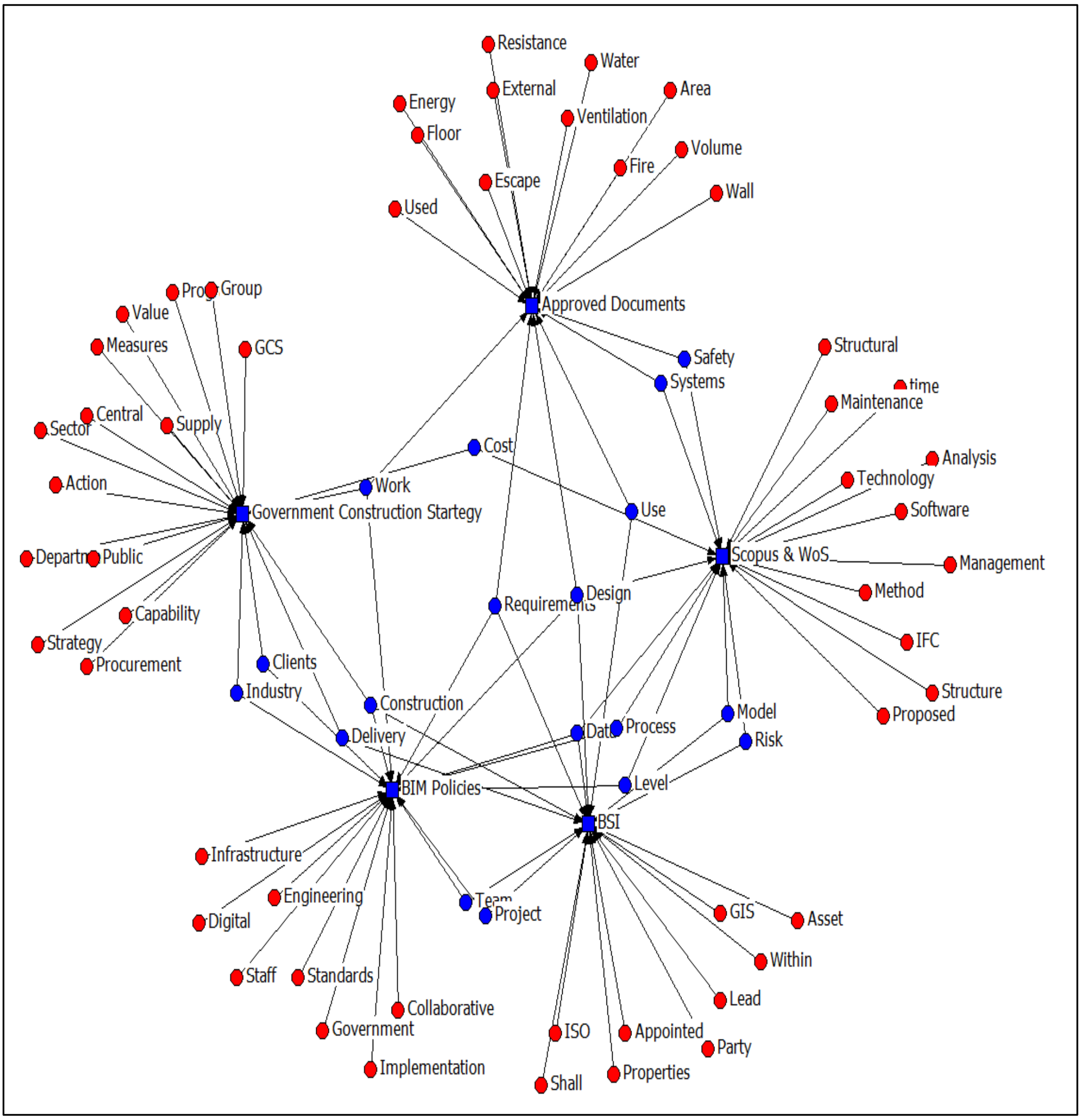


Figure 4.87: K-cores to identify the most influential nodes

The k-core analysis of the network of words in the experiments reveals that there are certain words that are highly connected across different experiments. The presence of these highly connected words suggests that these topics or concepts are commonly discussed and may play a crucial role in the discussions surrounding the experiments.

It can be seen that the most connected words between all experiments include “safety”, “systems”, “use”, “design”, “requirements”, “model”, “risk”, “process”, “level”, “data”, “cost”, “work”, “clients”, “industry”, “delivery”, “construction”, “team”, and "project". These words form the core structure of the network and indicate that they are central to the discussions in the experiments.

Furthermore, the k-core analysis can also help to identify the relationships between these central words and other less central words. This information can provide insights into the main themes and topics of discussion in the experiments and how they relate to one another. The k-core analysis provides valuable insights into the central structure and themes of the network of words in the experiments. This information can be used to further understand the relationships between different words and their role in the experiments.

4.11.6 Geodesic Distance

The calculation of geodesic distance in the network helps to understand the proximity of nodes within the network. The geodesic distance is defined as the shortest path between two nodes and the number of links between them. According to the data presented in the Table 4.42 below, it can be observed that a small percentage (0.019%) of the total pairwise combinations of nodes are only one step apart. This suggests that these nodes are closely connected and have a strong relationship within the network. On the other hand, a slightly smaller percentage (0.014%) of the pairwise combinations of nodes are two steps apart, indicating that these nodes are less connected. The table was created by calculating the frequency of pairwise combinations of nodes at each geodesic distance in the network, and then determining the proportion of each frequency relative to the total number of pairwise combinations.

Table 4.42: Nodes frequency

Geodesic Distance (Steps)		**Frequencies**	**Prop**
1	1	95	0.019
2	2	72	0.014

The Figure 4.88 provides a visual representation of the distribution of distances between nodes in the network. For a comprehensive examination of the complete-sized figure, please refer to Annex 3. This information can be useful in identifying clusters of nodes that are closely connected, as well as nodes that are more isolated from the rest of the network. The histogram can also be used to determine the average distance between nodes, which can provide insights into the overall structure of the network. By analysing the geodesic distances, the researcher can gain a better understanding of the relationships between nodes within the network.

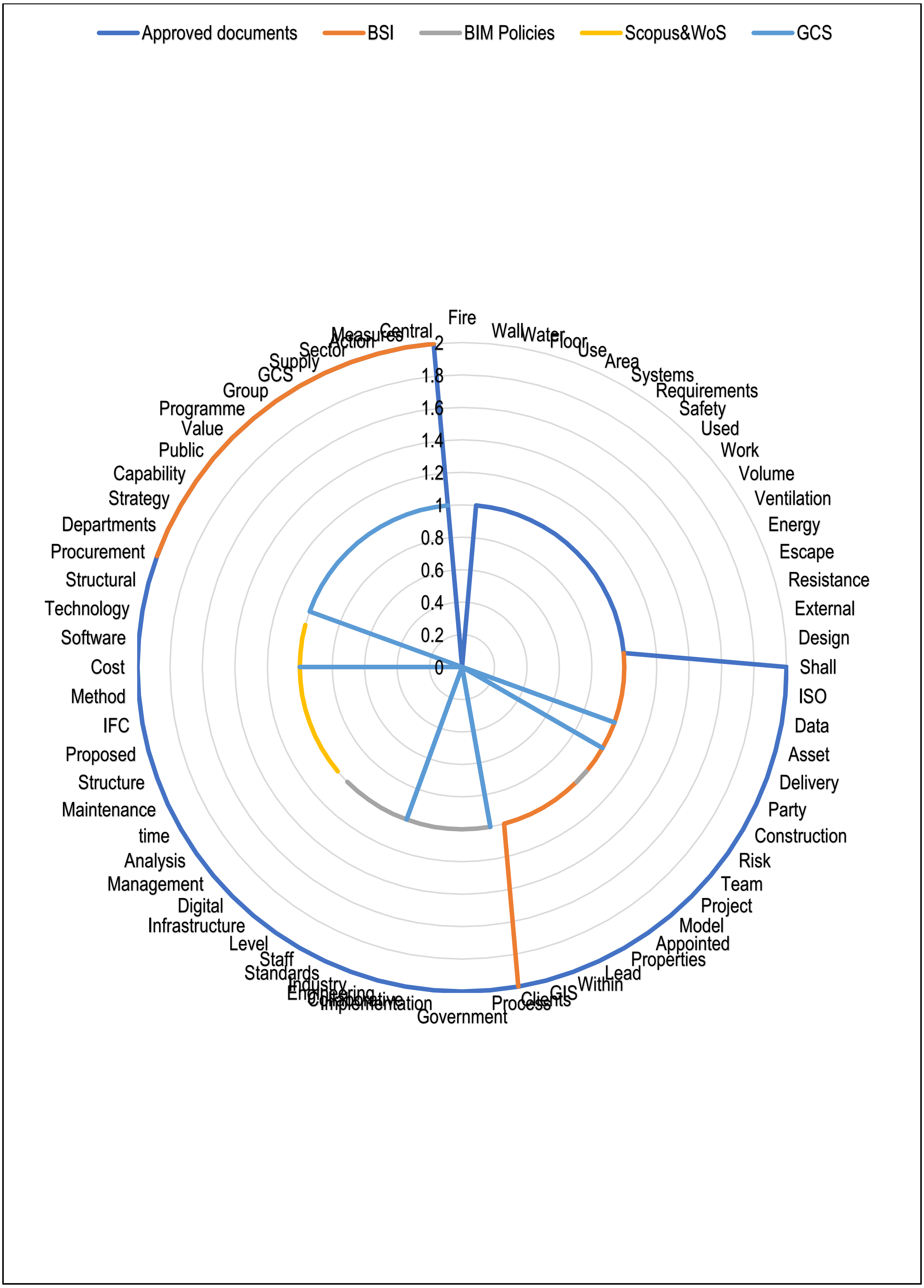


Figure 4.88: Distances of nodes in the network

4.11.7 Density

Additionally, the researcher also calculated the network density, which is a measure of the connectivity of the network. The density of a network can be defined as the ratio of the number of existing edges to the number of possible edges. It provides an insight into how many edges are present in the network compared to what could be possible. The network density can range from 0, indicating a completely disconnected network, to 1, indicating a completely connected network (Kanrak and Nguyen, 2022).

The density of the network was calculated and is presented in Table 4.43, which shows that the density of the network is 5.79%, which is calculated by dividing the number of actual connections or edges in the network by the total number of possible connections

Table 4.43: Network density

Density (matrix average)	0.0579
Standard deviation	0.1624

Despite the modest level of density, the results of the experiments indicate that the connections in the network are not particularly strong. This is due to the fact that the researcher carried out the SNA only on the 20 most frequent words from each experiment. By limiting the number of words, the likelihood of collisions, inaccuracies, and errors is reduced, leading to more accurate results of the analysis.

4.11.8 Degree Centrality

Degree centrality is a measure of the importance of a node in a graph based on the number of connections it has to other nodes. It is a simple and widely used measure of centrality in a graph. In a directed graph, there are two types of degree centrality: in-degree and out-degree. The in-degree centrality of a node is the number of incoming edges to that node, i.e., the number of nodes that are connected to it. The out-degree centrality of a node is the number of outgoing edges from that node, i.e., the number of nodes that it is connected to. In an undirected graph, there is only one type of degree centrality, which is simply the total number of edges incident to a node (Zeng et al., 2023).

Degree centrality, as expounded by Zeng et al. (2023), is a common metric used to assess the relative importance of a node in a graph. It is defined as the number of edges that are incident to a node and serves as a simple indicator of the node's connectivity within the network. While degree centrality can provide insights into the number of connections a node has, it should be used with caution as it does not take into consideration the structural context of the network, such as the distances between nodes or the weights of the edges. As a result, it may not always accurately reflect the true importance of a node in the network and can be biased towards nodes with a high number of connections, regardless of their actual influence within the network.

In order to evaluate the performance of the top 20 frequent words from the experimental studies, centrality measures including degree centrality, closeness centrality, and "betweenness" centrality were employed. The results based on degree centrality revealed that words such as "fire" (17), "wall" (20), "water" (20), "floor" (21), and "use" (22) had the highest co-authorship with other words. These were followed by "requirements" (3), "safety" (2), "work" (3), "design" (4), "data" (3), "delivery" (2), "construction" (3), "risk" (2), "team" (2), "project" (2), "model" (2), "clients" (2), "process" (2), "industry" (2), "level" (2), and "cost" (2). All other words had a degree centrality of one.

According to the betweenness centrality, words such as "use" (988), "fire" (877), "floor" (799), "wall" (791), and "water" (678) were found to be well positioned in the network and had a higher likelihood of being a part of the shortest path between two words. These words also play a crucial role in controlling the flow of information within the network. This is followed by words such as "cost" (100), "design and data" (78.594), "construction" (49.92), "work and clients" (33.041), "risk and model" (30.886), "level and process" (30.829), and "team and project" (16.879).

Regarding the closeness centrality measure, words such as "government", "implementation", "collaborative", "engineering", "standards", "staff", "infrastructure", and "design" (230) have the shortest distances to other nodes in the network. This high closeness centrality index represents their effectiveness and central role in the distribution of information among other nodes in the network (as seen in Table 4.34). This is followed by words such as "management", "analysis", "time", "maintenance", "structure", "proposed", "IFC", "method", "software", "technology", and "structural" (228). The network of these measures is also depicted in Figures 4.89, 4.90, and 4.91 and Tables 4.44 and 4.45.

Table 4.44: Frequencies of network words (highest to lowest)

	Words	Frequencies	Centrality		
			Degree	**Closeness**	**Betweenness**
			Frequencies	Frequencies	Frequencies
1	Fire	2840	17	125	877.002
2	Wall	2322	20	143	791.585
3	Water	1474	20	160	678.835
4	Floor	1331	21	158	799.975
5	Use	1200	22	141	988.773
6	Area	1065	1	195	0
7	Systems	1877	2	177	0
8	Requirements	955	3	165	16.879
9	Safety	937	2	177	0
10	Used	923	1	195	0
11	Work	916	3	161	33.041
12	Volume	914	1	195	0
13	Ventilation	911	1	195	0
14	Energy	826	1	195	0
15	Escape	822	1	195	0
16	Resistance	755	1	195	0
17	External	750	1	195	0
18	Design	737	4	152	78.594
19	Shall	685	1	213	0
20	ISO	651	1	213	0
21	Data	526	3	161	78.594
22	Asset	449	1	213	0
23	Delivery	399	2	194	0
24	Party	390	1	213	0
25	Construction	355	3	172	49.92
26	Risk	584	2	182	30.886
27	Team	296	2	188	16.879
28	Project	296	2	188	16.879
29	Model	282	2	182	30.886
30	Appointed	277	1	213	0

Table 4.44: Frequencies of network words (highest to lowest)

31	Properties	213	1	213	0
32	Lead	212	1	213	0
33	Within	205	1	213	0
34	GIS	205	1	213	0
35	Clients	56	2	184	33.041
36	Process	53	2	198	30.829
37	Government	52	1	230	0
38	Implementation	47	1	230	0
39	Collaborative	45	1	230	0
40	Engineering	44	1	230	0
41	Industry	43	2	184	33.041
42	Standards	39	1	230	0
43	Staff	39	1	230	0
44	Level	39	2	198	30.829
45	Infrastructure	39	1	230	0
46	Digital	37	1	230	0
47	Management	2929	1	228	0
48	Analysis	1608	1	228	0
49	time	1472	1	228	0
50	Maintenance	1429	1	228	0
51	Structure	1374	1	228	0
52	Proposed	1328	1	228	0
53	IFC	1325	1	228	0
54	Method	1313	1	228	0
55	Cost	1276	2	177	100.534
56	Software	1212	1	228	0
57	Technology	1203	1	228	0
58	Structural	1151	1	228	0
59	Procurement	83	1	211	0
60	Departments	71	1	211	0
61	Strategy	68	1	211	0
62	Capability	64	1	211	0
63	Public	59	1	211	0

Table 4.44: Frequencies of network words (highest to lowest)

64	Value	52	1	211	0
65	Programme	47	1	211	0
66	Group	47	1	211	0
67	GCS	46	1	211	0
68	Supply	45	1	211	0
69	Sector	45	1	211	0
70	Action	44	1	211	0
71	Measures	41	1	211	0
72	Central	41	1	211	0

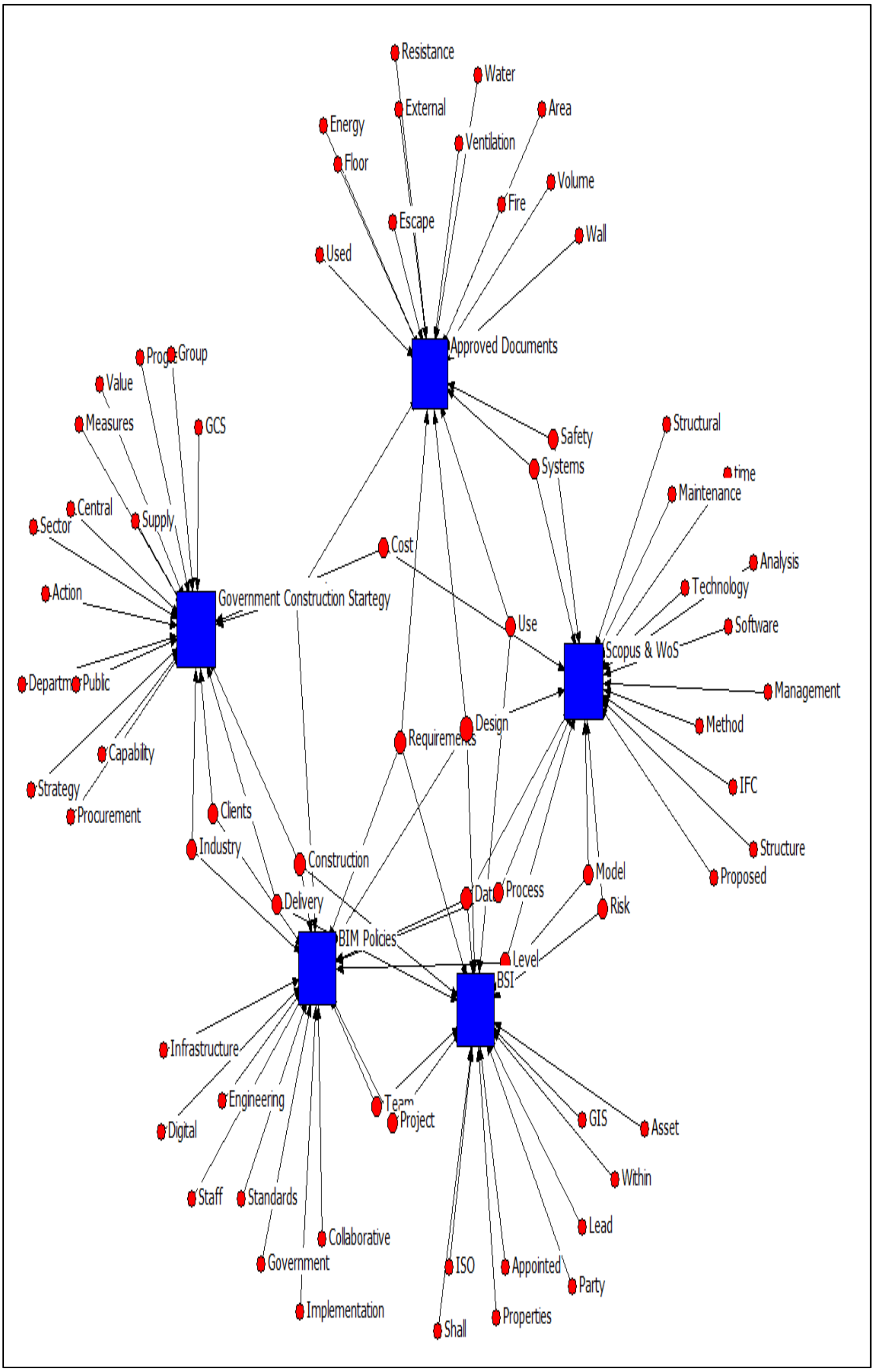


Figure 4.89: Frequency of network words

Table 4.45: Degree centrality

Rank	Words	Degree	nDegree		Rank	Words	Degree	nDegree
1	Fire	17	0.239		37	Government	1	0.014
2	Wall	20	0.282		38	Implementation	1	0.014
3	Water	20	0.282		39	Collaborative	1	0.014
4	Floor	21	0.296		40	Engineering	1	0.014
5	Use	22	0.31		41	Industry	2	0.028
6	Area	1	0.014		42	Standards	1	0.014
7	Systems	2	0.028		43	Staff	1	0.014
8	Requirements	3	0.042		44	Level	2	0.028
9	Safety	2	0.028		45	Infrastructure	1	0.014
10	Used	1	0.014		46	Digital	1	0.014
11	Work	3	0.042		47	Management	1	0.014
12	Volume	1	0.014		48	Analysis	1	0.014
13	Ventilation	1	0.014		49	time	1	0.014
14	Energy	1	0.014		50	Maintenance	1	0.014
15	Escape	1	0.014		51	Structure	1	0.014
16	Resistance	1	0.014		52	Proposed	1	0.014
17	External	1	0.014		53	IFC	1	0.014
18	Design	4	0.056		54	Method	1	0.014
19	Shall	1	0.014		55	Cost	2	0.028
20	ISO	1	0.014		56	Software	1	0.014
21	Data	3	0.042		57	Technology	1	0.014
22	Asset	1	0.014		58	Structural	1	0.014
23	Delivery	2	0.028		59	Procurement	1	0.014
24	Party	1	0.014		60	Departments	1	0.014
25	Construction	3	0.042		61	Strategy	1	0.014
26	Risk	2	0.028		62	Capability	1	0.014
27	Team	2	0.028		63	Public	1	0.014
28	Project	2	0.028		64	Value	1	0.014
29	Model	2	0.028		65	Programme	1	0.014
30	Appointed	1	0.014		66	Group	1	0.014
31	Properties	1	0.014		67	GCS	1	0.014
32	Lead	1	0.014		68	Supply	1	0.014
33	Within	1	0.014		69	Sector	1	0.014
34	GIS	1	0.014		70	Action	1	0.014
35	Clients	2	0.028		71	Measures	1	0.014
36	Process	2	0.028		72	Central	1	0.014

The data is presenting the results of a centrality analysis of a network based on the 20 most frequent words in BIM experiments. The centrality measures were degree centrality and normalized degree centrality.

According to the degree centrality measure, the five words with the highest centrality are “fire”, “wall”, “water”, “floor”, and “use”. This suggests that these words are the most interconnected and connected to the largest number of other words in the network. On the other hand, the words with the lowest degree centrality are “area”, “volume”, “ventilation”, and “ISO”. These words have the smallest number of connections in the network.

The normalized degree centrality measure provides a similar result, with the top five words being “fire”, “wall”, “water”, “floor”, and “use”. This measure normalizes the degree centrality scores by the number of nodes in the network, making it easier to compare centrality scores between networks of different sizes. The words with the lowest normalized degree centrality are “government”, “implementation”, “collaborative”, and “engineering”.

The high degree centralization value (0.2805) and eigenvector centralization value (58.395) indicate a highly centralized network. This suggests that the top 20 most frequent words in BIM experiments are strongly correlated, meaning that they are highly interconnected and likely to appear together in the data.

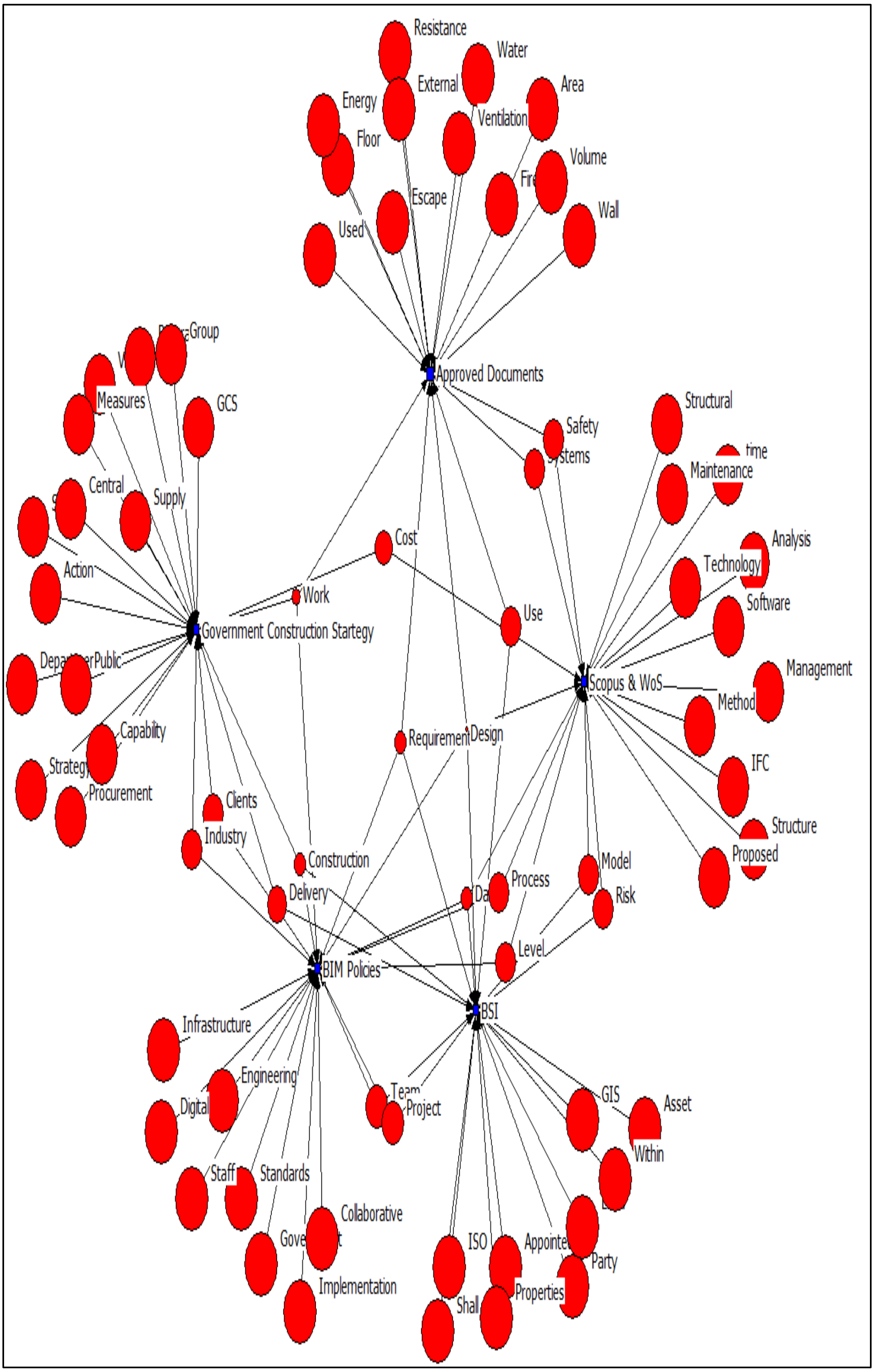


Figure 4.90: Frequency of network words

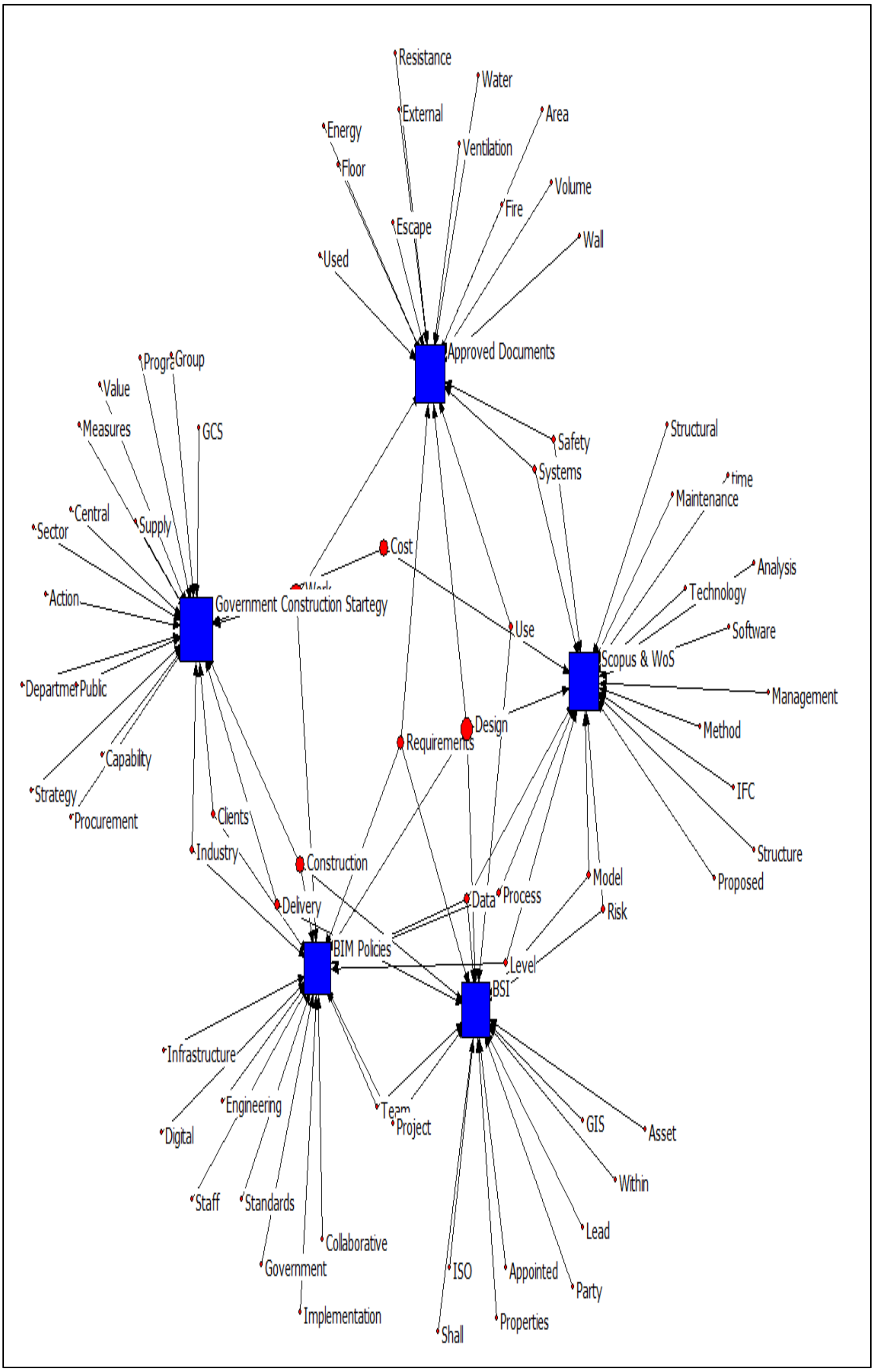


Figure 4.91: Frequency of network words

4.11.9 Eigenvector Centrality

In SNA, eigenvectors are utilized to evaluate the significance of nodes in a network. The concept is that a node that has many connections to other important nodes is considered important itself. One method to determine the importance of a node is by using eigenvector centrality. This measurement assesses the impact of a node in a network by assigning relative scores to all nodes based on the idea that connections to high-scoring nodes carry more weight in determining the node's score than equal connections to low-scoring nodes (Kiburu et al., 2023).

For this study, a social network was created whereby the top 20 most frequent words served as the nodes, and their relationships were represented by edges. The eigenvector centrality was used to measure the importance of each word in the network. A word that is connected to many other important words will have a high eigenvector centrality score, while a word with only a few connections that are not well-connected will have a low eigenvector centrality score. Eigenvector centrality is often used in combination with other centrality measures, such as degree centrality and betweenness centrality, to gain a more comprehensive understanding of the significance of each node in the network (Kiburu et al., 2023). Figure 4.92 displays the eigenvector centrality of the top 20 most frequent words, with the words with the highest eigenvector centrality being "design", "requirements", "work", "construction", "data", and "cost". For more information, see Table 4.46.

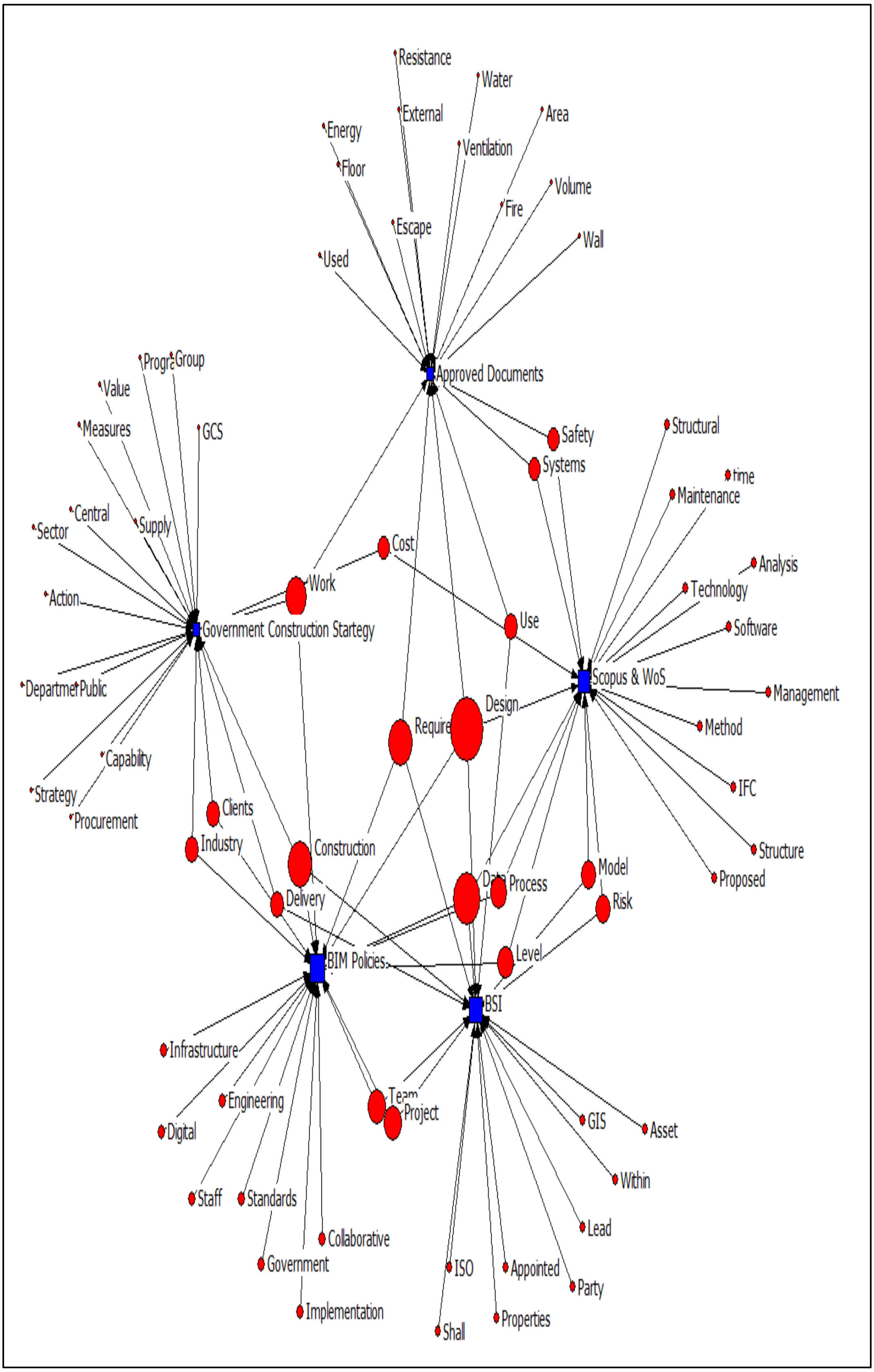


Figure 4.92: Eigenvector centrality

Table 4.46: Eigenvector centrality

Rank	Words	Eigenvector		Rank	Word	Eigenvector
1	Fire	0.446		37	Government	0.049
2	Wall	0.353		38	Implementation	0.049
3	Water	0.332		39	Collaborative	0.049
4	Floor	0.287		40	Engineering	0.049
5	Use	0.317		41	Industry	0.096
6	Area	0.066		42	Standards	0.049
7	Systems	0.108		43	Staff	0.049
8	Requirements	0.167		44	Level	0.091
9	Safety	0.108		45	Infrastructure	0.049
10	Used	0.066		46	Digital	0.049
11	Work	0.162		47	Management	0.042
12	Volume	0.066		48	Analysis	0.042
13	Ventilation	0.066		49	time	0.042
14	Energy	0.066		50	Maintenance	0.042
15	Escape	0.066		51	Structure	0.042
16	Resistance	0.066		52	Proposed	0.042
17	External	0.066		53	IFC	0.042
18	Design	0.209		54	Method	0.042
19	Shall	0.052		55	Cost	0.089
20	ISO	0.052		56	Software	0.042
21	Data	0.144		57	Technology	0.042
22	Asset	0.052		58	Structural	0.042
23	Delivery	0.099		59	Procurement	0.047
24	Party	0.052		60	Departments	0.047
25	Construction	0.148		61	Strategy	0.047
26	Risk	0.095		62	Capability	0.047
27	Team	0.101		63	Public	0.047
28	Project	0.101		64	Value	0.047
29	Model	0.095		65	Programme	0.047
30	Appointed	0.052		66	Group	0.047
31	Properties	0.052		67	GCS	0.047
32	Lead	0.052		68	Supply	0.047
33	Within	0.052		69	Sector	0.047
34	GIS	0.052		70	Action	0.047
35	Clients	0.096		71	Measures	0.047
36	Process	0.091		72	Central	0.047

After conducting the analysis, the researcher presents a graph based on the most significant components, as depicted in Figure 4.93. Upon inspection of the graph, it becomes evident that all of the results from the experiments are closely linked. This indicates that the aspects identified through the study are widely discussed in the context of BIM and its potential for improving performance in large-scale project delivery. This observation is noteworthy as it highlights the importance of these topics within the BIM community, which could potentially inform future research directions in this field. Additionally, this result could also be used to guide the prioritization of BIM implementation efforts in the industry, as it identifies the areas of concern that are most commonly discussed and could potentially have the greatest impact on project performance.

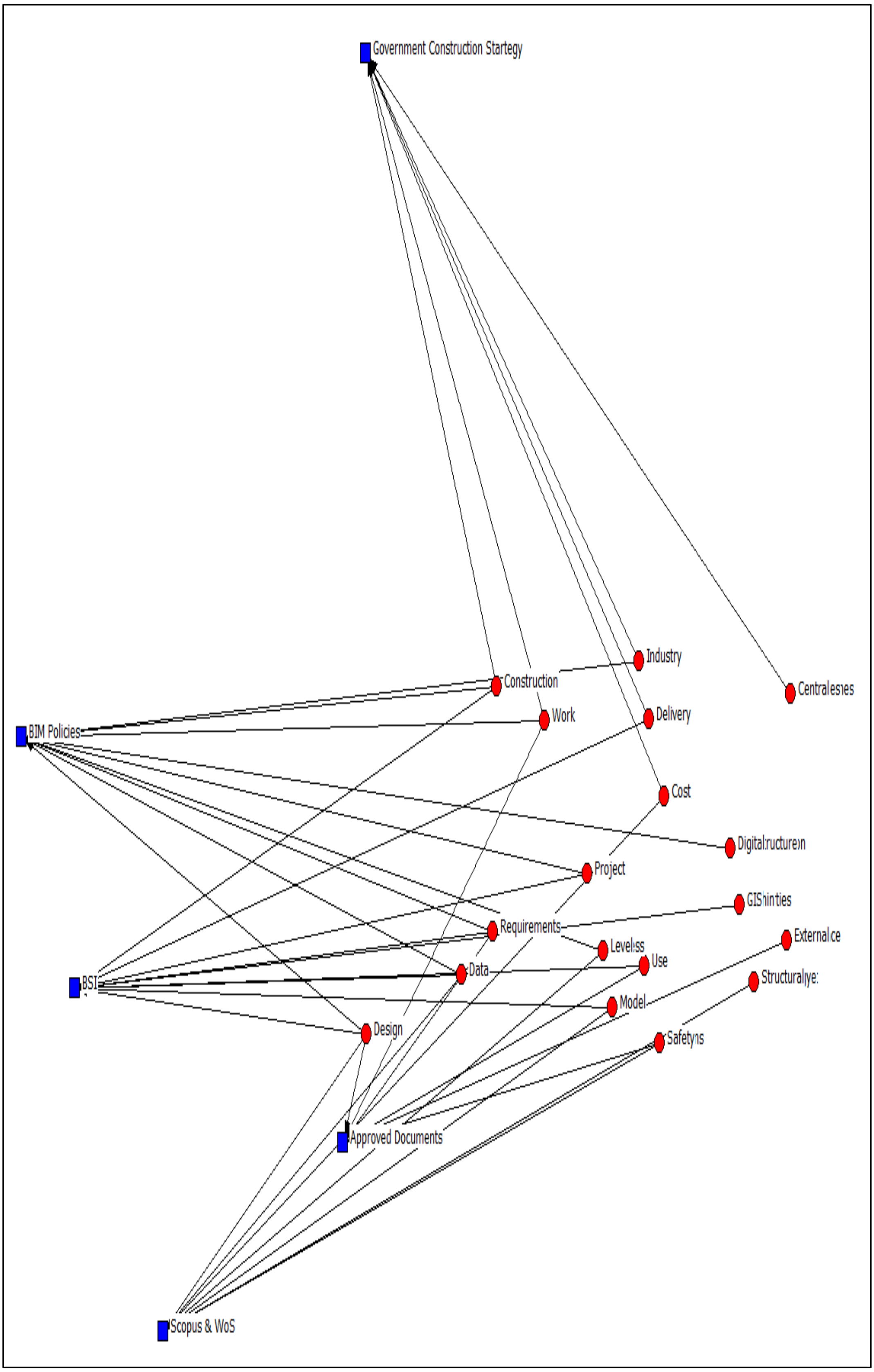


Figure 4.93: Output depicting interconnectedness of the experiments

4.12 Summary

This chapter presents the experimental study and validation of the BIMTAPE Framework. The chapter begins with an introduction highlighting the challenges associated with text analysis. It then proceeds to explain the development of the BIMTAPE Framework, which includes steps such as BIM text collection, pre-study BIM business understanding, BIM data modelling (TM), and BIM model validation. The chapter describes a series of experiments undertaken to achieve these tasks.

Experiment 1 involves analysing academic articles related to BIM using the BIMTAPE Framework. It provides details about the body of text analysed and the text analysis process employed.

Experiment 2 focuses on analysing technical files from case projects using the BIMTAPE Framework. The chapter discusses the tools used for the study, the body of text analysed, and the process of TM.

Experiments 3-6 involve analysing different types of documentation, including regulation documentations, British Standards, national policy papers, and enterprise policy papers.

The chapter provides an introduction to each experiment, describes the body of text collected, and outlines the TM process employed for each experiment. The chapter also delves into BIM model validation, introducing SNA as a method. It discusses building relationship graphs, creating a network of the 20 most frequent words, and utilizing a two-mode network for analysis. Overall, this chapter presents a comprehensive overview of the experimental study conducted using the BIMTAPE Framework. It highlights the different types of text analysed and the processes employed, ultimately leading to valuable insights for performance enhancement into BIM pervasive major project delivery.

Chapter 5
Justifying BIMTAPE Framework

5.1 Introduction

The strategic integration of TM methodologies holds significant promise for enhancing performance in BIM pervasive major project delivery. TM's capacity to extract insights from extensive textual data opens doors to competitive advantages, improved decision-making, and optimized project processes. While its application introduces complexities in data analysis and interpretation, the benefits of uncovering trends and patterns are substantial. This chapter explores the symbiotic relationship between TM and project performance, emphasizing the role of BIM models in guiding insights. Through the lens of TM, this chapter unveils how data-driven insights can transform the landscape of major construction projects.

5.2 Leveraging TM for Enhanced Performance

TM methodologies, when correctly applied, can lead to improved performance in BIM pervasive major project delivery. By using TM, firms can gain a competitive advantage by extracting meaningful insights from large amounts of data, helping to identify potential problems and optimize the construction process. This leads to better communication and collaboration between stakeholders, resulting in more accurate and efficient project delivery. By utilizing TM, companies can gain a competitive advantage by improving their ability to extract insights from large amounts of data (Ittoo et al., 2016), which can help in the identification of potential problems and optimization of the construction process. This leads to improved communication and collaboration between stakeholders, resulting in more efficient and accurate project delivery.

However, the implementation of TM in BIM pervasive major project delivery is not without challenges. One of the main limitations is the complexity of analysing and processing data in the construction industry (Love et al., 2020; Zhou et al., 2013). This complexity is further compounded by the unstructured nature of much of the data involved, making it difficult to integrate information effectively across various stages of construction projects (Liu and Issa, 2016). The systems of text analytics used in

BIM need to have domain knowledge expertise in order to accurately capture industry-specific keywords. This requires companies to invest in the development of specialized software and the training of personnel, in order to ensure that the systems are able to effectively process and analyse the data generated during a project.

TM has revealed the significant role that BIM models play in the success of BIM pervasive major project delivery. By providing a governance structure, the BIM model helps to identify problems and provides the means for communication and collaboration between stakeholders.

Research has demonstrated that the use of Industry Foundation Classes (IFC) enhances the utilization and adoption of BIM in the construction industry, as they provide a platform for improved data communication and iterative workflows. The analysis of data through TM and co-occurrence keywords has revealed a correlation between energy management, performance, and BIM. Model view definitions, which are IFC subsets (Ghaffarianhoseini et al., 2017), facilitate the communication and optimization of construction processes, leading to improved project delivery outcomes.

5.3 Use of BIM in Major Construction Projects

BIM is a tool that plays a significant role in supporting the management of the major construction projects, ensuring that it is possible to manage cost overruns and revenue stream of the major construction projects without hampering economic growth. The policies, technologies and associated processes of BIM support communication, analysis, and production of the constructive models that enhance construction stakeholders' engagement in collaborative, operation and building of the facility. Furthermore, BIM ensures that major construction project delivery has a team that is focused on realization of project results. The team is under leadership that supports effective communication among team members to realize positive outcomes in the construction project (Baiden and Price, 2011). BIM also ensures that team productivity in the construction sector is improved, particularly in terms of time and cost efficiency (Bryde et al., 2013). PMs are also able to achieve success since BIM helps them in coordination and communication in the construction projects.

Stakeholders and other players who are involved in the construction sector have the mandate of ensuring that they focus on adopting and implementing modern levels of BIM technologies. The approach will result in improving productivity of the workforce and realization of better performance for the organization while improving on operational efficiency. Furthermore, BIM supports decision-making within construction projects, and ensures that stakeholders have a comprehensive approach to realization of positive results in construction works. Such is enhanced through the establishment of strong stakeholder relationships and needs that lead to the improvement of performance and realization of success in construction projects.

BIM also provides technical solutions in cost management, quality and safety management, technical management, and field management (Pan et al., 2021). The technological tool also facilitates in detection of clash analysis in the phase of project optimization and design for quality, schedule, and cost. Trend forecasting and hotspots application of BIM is supported by patent analysis, text clustering, and SNA. Unfortunately, BIM tools often experience challenges within the construction sector that are linked to issues of intellectual property, cultural differences, ownership, and misunderstandings (Xiong et al., 2017; Wang et al., 2020). Such situation contributes to problems that include data generation errors, loss, discrepancies, and incorrect or incomplete data. However, the technical features and capability of BIM ascertains that these aspects are identified and addressed to realize positive results.

5.4 Construction Projects Stakeholders

Stakeholders who are involved in construction projects often create challenges to project progress. The situation occurs when these stakeholders fail to establish relationships and an element of interconnectedness. Hence, projects tend to stall and take longer durations than expected because of the lack of effective interactions among the stakeholders (Loosemore, 2015; Othman and Nadim, 2010). Weak relationships among stakeholders limit the effective utilisation of construction resources. The lack of interaction among stakeholders could also hinder awareness of the BIM technology, which indicates that the implementation of such a tool does not receive positive support within the construction projects (Bryde et al., 2013; Cao et al., 2017).

TM is an approach that is associated with several challenges, especially within the construction industry. The challenges are attributed to the complexity of capturing the data and analysing the same. As such, the data sources are different, and the storage and retrieval of the same are different, indicating that unique approaches have to be adopted in the process of analysing the data (Lu et al., 2020; O'Connor and Yang, 2004). The heterogeneity of the data also subjects it to different forms and ways of encoding. Therefore, domain knowledge expertise is needed when focusing on text analytics systems for the BIM pervasive major project delivery. The approach is needed to ensure that there is effective and correct capturing of the used keywords (Merschbrock and Munkvold, 2015; Ittoo et al., 2016).

The new technologies that support data analysis enhance value delivery through support of real-time decision-making. The applications that focus on handling and management of data are expected to generate outcomes, as well as deliver real-time results without incurring any disruptions (Azhar, 2011; Bourreau et al., 2018). The rate of data generation and utilization in social networks needs to be improved to produce data, which is unique and not challenged (Hassani et al., 2020).

5.5 BIMTAPE Framework

To support the implementation of BIM in construction projects, several frameworks and methodologies have been developed to help in the analysis of BIM-related texts. One of the most prominent frameworks is the TAPE, which (as explained previously) was specifically designed to support the analysis and interpretation of texts in BIM. The developed BIMTAPE Framework focuses on the analysis of text purpose, context, relevance, and quality of the contained information. This framework provides a structure for the analysis of BIM-related texts, including standards, contracts, technical documents, and specifications.

The framework is implemented through the use of a software tool, which transforms the formal process of development into a project format that the tool can work with (Kuhrmann et al., 2014). BIMTAPE Framework has been shown to be highly effective in the analysis of the message that is provided in the text. The framework has helped to understand the potential of BIM in enhancing performance of major construction projects.

5.6 BIMTAPE Framework Contribution and Impact

By employing the BIMTAPE Framework, stakeholders can expect significant improvements in the performance of BIM pervasive major project delivery. The BIMTAPE Framework enhances the understanding of BIM-related texts by analysing their purpose, context, relevance, and quality of information. By applying this framework, stakeholders can extract valuable insights from BIM-related texts. This deeper understanding enables stakeholders to make informed decisions and better manage the complexities of major construction project delivery.

The BIMTAPE Framework also enables the identification of keywords and their functions. By conducting a comprehensive analysis of these keywords, stakeholders can gain valuable insights into the critical aspects of major construction project delivery. For example, they can identify KPIs, potential risks, areas for improvement, and effective communication strategies. This knowledge empowers stakeholders to optimize project performance, mitigate risks, and enhance collaboration among project participants.

Moreover, the BIMTAPE Framework facilitates the identification of patterns and relationships. By using advanced TM techniques, such as co-occurrence analysis and SNA, the framework allows stakeholders to visualize and analyse the connections between various concepts and stakeholders.

By utilizing the BIMTAPE Framework, stakeholders can address the identified research gap to improve performance in major construction project delivery. The practical implications and potential benefits of the BIMTAPE Framework are numerous. For policy makers and government bodies, the framework can inform the development of regulations, guidelines, and standards that promote the effective use of BIM in major construction project delivery. It enables them to leverage the vast amount of information contained within BIM-related texts to ensure compliance, optimize processes, and drive industry-wide improvements.

Within the construction industry, companies and professionals can benefit from the BIMTAPE Framework by enhancing their project management strategies and optimizing resource allocation. Ultimately, the application of the BIMTAPE

Framework contributes to the advancement of BIM pervasive major project delivery by leveraging the power of text analysis.

5.7 Theory of Inventive Problem Solving (TRIZ)

The TRIZ is a valuable tool for analysing and forecasting problems. It provides a structured approach to innovation, in which problem assumptions are analysed in a way that makes them solvable (ASQ, 2022). As a result, it plays a crucial role in conducting extensive systematic literature review s and evaluating articles. The research outcomes presented in this thesis were based on the application of the NSP -guided method of EBL and the authors' observations from relevant research and practices in major construction projects.

The application of the NSP approach led to the development of the BIMTAPE Framework, which focuses on text analysis in the context of BIM. The NSP approach, in combination with the BIMTAPE Framework and TM, provides a comprehensive solution for the analysis and optimization of BIM pervasive major project delivery.

5.8 Summary

In this chapter, the BIM-oriented insights gained from the research are summarized, focusing on TM, the use of BIM in major construction projects, the TAPE framework, and the NSP. These have provided valuable insights into BIM pervasive major project delivery, leading to enhanced understanding and potential improvements in the construction industry.

TM has emerged as a powerful tool for extracting meaningful insights from large amounts of data. The use of BIM in major construction projects has been shown to be crucial for success. BIM facilitates effective communication, coordination, and collaboration among project stakeholders, leading to improved accuracy, efficiency, and overall project performance (Smith, 2020). It provides technical solutions for various aspects such as cost management, quality control, safety management, and field management, enabling the optimization of construction processes and the achievement of positive project outcomes.

The TAPE framework focuses on understanding the purpose, context, relevance, and quality of information contained in the texts, enabling stakeholders to extract key insights and make informed decisions.

The NSP provides a systematic approach to innovation and problem-solving. Through NSP-guided methods like EBL, the research outcomes have been enriched, allowing for a comprehensive analysis of BIM pervasive major project delivery and the identification of potential improvements. NSP aids in the systematic examination of problem assumptions, leading to solvable challenges and innovative solutions in the construction industry.

Chapter 6
Findings and Discussion

6.1 Introduction

This chapter presents the findings of the experimentation conducted in this study and compares them with the findings from the systematic literature review section. The purpose is to examine the alignment and consistency between the empirical results obtained through experimentation and the theoretical insights derived from the systematic literature review. The chapter begins with a brief recapitulation of the key findings from the experimentation and systematic literature review, followed by a discussion of their similarities, differences, and implications. The aim is to provide a comprehensive analysis of the research findings and their implications for BIM pervasive major project delivery.

6.2 Experimentation Findings

The experimentation in this study involved various approaches, including TM, web scraping, clustering, topic modelling, and analysis of term-document matrices. The results from each experiment are summarized below.

6.2.1 Summary of Outcomes

Experiment 1: The study found that BIM publications and research trends are emerging from Asia, the USA, Europe, New Zealand, and Australia. African countries were found to have limited presence in BIM research, indicating challenges in BIM adoption. However, the construction sector globally is witnessing the adoption of BIM due to its potential for environmentally friendly and efficient construction practices (Azhar, 2011). This aligns with the broader goals of the BIMTAPE Framework and highlights the need for a global perspective in shaping BIM strategies for major project delivery.

Relating these findings to the insights gained from the BIMKG, the global adoption trend resonates with the interconnected nature of BIM elements. The BIMKG, by identifying key terms like Data, Design, and System and their strategic implications, provides a foundation for understanding how BIM practices contribute to efficiency and sustainability objectives in major project delivery.

In integrating the BIMKG insights with Experiment 1's findings, it becomes apparent that a comprehensive understanding of BIM elements and their relationships is pivotal for addressing challenges and promoting adoption worldwide. The BIMTAPE Framework, with its emphasis on performance enhancement in major project delivery, can benefit from these collective insights by adapting strategies to be contextually aware and globally relevant.

Therefore, the findings from Experiment 1 and the BIMKG collectively emphasise the importance of a holistic and globally informed approach to BIM implementation. This approach acknowledges regional variations in research engagement while recognising the overarching global trend towards adopting BIM for its transformative potential in sustainable and efficient construction practices.

Experiment 2: Web scraping and TM techniques were employed to study the use of BIM in enhancing performance in major construction project delivery. The analysis revealed frequent words related to performance, efficiency, time, quality, and cost. The visualisation of topics generated by the LDA model highlighted the relevance of BIM in different stages of construction and performance enhancement.

The observed emphasis on key performance indicators such as efficiency, time, quality, and cost aligns with the goals of the BIMTAPE Framework, which seeks to improve overall performance in major project delivery. Experiment 2's findings resonate with the interconnectedness highlighted by the BIMKG, further reinforcing the importance of understanding the relationships between BIM key terms for effective implementation.

In parallel, the BIMKG, as illustrated in Figure 4.25, visually represents intricate relationships, and influences among different facets of BIM. The connections between topics such as Tunnels and Ground, as indicated by edges in the knowledge graph, signify relationships between these specific areas. This graphical representation significantly aids in comprehending the complex interplay between BIM key terms, providing a visual roadmap for understanding their interconnectedness.

Experiment 2's emphasis on performance aligns seamlessly with the BIMKG's depiction of relationships, collectively supporting the goals of the BIMTAPE Framework. This synthesis of findings reinforces the notion that a comprehensive understanding of BIM's impact on performance involves recognising not only individual performance indicators but also their interdependencies across various BIM domains.

The combination of Experiment 2's textual analysis and the visual representation offered by the BIMKG contributes to a holistic comprehension of how BIM elements interact and influence performance in major construction projects. These findings provide practitioners with a strategic perspective, aligning with the BIMTAPE Framework's objectives and fostering continual improvement in BIM implementation.

Experiment 3: The analysis of frequently used words and their relationships in the corpus demonstrated the importance of visualising top frequent terms and applying feature selection techniques. The associations between terms provided insights into the relationships between different concepts. Topic modelling using LDA allowed for the identification of main themes and topics in the text data.

Experiment 4: The analysis of BSI documents using term-document matrices, clustering, and topic modelling techniques provided insights into word relationships, groupings, and main themes. The clusters and topics revealed the relevance of BIM in various aspects of major construction project delivery, such as procurement, cost optimization, and spatial coordination.

Experiment 5: The analysis of BIM policy documents highlighted the coverage and focus areas of these policies. The most frequent terms indicated an emphasis on collaboration, government involvement, and infrastructure. The clustering analysis grouped terms related to implementation strategies, collaboration practices, and project management aspects.

Experiment 6: The analysis of GCS documents revealed an emphasis on construction, cost, procurement, industry, client, and strategy. The topics identified in the GCS documents focused on optimizing procurement, cost, and digital capability. The findings emphasized the importance of these aspects in achieving successful outcomes in BIM pervasive major project delivery.

The insights derived from the BIMKG unveil profound connections between various thematic strands in Experiments 3, 4, 5, and 6, enriching the understanding of the complex relationships within the Periodic Table of BIM. Notably, the identified relationships between TS5 and TS3, as illustrated in Figure 4.70 within the BIMKG, encompassing terms such as Work, Design, and Requirements, provide a nuanced perspective on the robust interplay between BIM documentation and the policies guiding BIM implementation within organisations. This alignment underscores the convergence of work processes, design principles, and specified requirements, signifying a harmonious integration between technical specifications and organisational BIM policies, as outlined in the Periodic Table of BIM.

Likewise, the correlation uncovered between TS4 and TS3, focusing on terms like Use, Design, and Requirements, suggests a synchronisation between British Standards (TS4) and fundamental concepts within BIM documentation (TS3). This alignment becomes pivotal in maintaining adherence to established standards and ensuring the seamless integration of BIM processes within the broader industry landscape, aligning with the principles of the Periodic Table of BIM.

The interconnected terms such as Team, Project, Data, Construction, Design, and Requirements in the connection between TS4 and TS5, as revealed by the BIMKG, unveil a substantial overlap between British Standards and BIM policies at companies. This alignment indicates that the standards outlined in TS4 significantly influence the shaping of team dynamics, project management, data handling, construction processes, and design aspects as delineated in BIM policies, providing a tangible link to the BIMTAPE Framework.

In a parallel vein, the connection between TS4 and TS6, particularly involving terms like Delivery and Construction, as depicted in the BIMKG, signifies that the standards articulated in British Standards (TS4) carry apparent implications for the delivery and construction phases of major projects. This connection underscores the influential role of standards in shaping project delivery processes on a national scale, aligning with the broader objectives of the BIMTAPE Framework.

The relationship observed between terms such as Working, Industry, and Construction in the connection between TS5 and TS6, illuminated by the BIMKG, underscores a synergistic alignment between organisational BIM policies and broader industry-related policies outlined in national policy papers. This connection emphasises the pivotal role of aligning company practices with industry-wide strategies in the effective implementation of BIM, a concept foundational to the BIMTAPE Framework.

6.2.2 Implications

The findings from the experiments mentioned above, which employed TM techniques, provide valuable insights, and can help improve performance in BIM pervasive major project delivery.

Experiment 1: Findings from experiment 1 can guide stakeholders to learn from successful implementations, best practices, and innovative approaches. This cross-regional knowledge exchange can contribute to enhancing performance by leveraging the collective wisdom of different regions.

Experiment 2: Findings can help project stakeholders to prioritize their efforts, allocate resources effectively, and develop strategies to optimize performance in these areas. The visualisation of topics further aids in identifying the specific stages of construction where BIM can have the most significant impact on performance enhancement, enabling targeted interventions.

Experiment 3: Insight derived from this experiment can help stakeholders understand the interconnectedness of various project elements and identify dependencies. By recognizing these relationships, project teams can enhance communication, collaboration, and coordination among stakeholders. This, in turn, improves the overall efficiency, effectiveness, and quality of major construction project delivery.

Experiment 4: Findings from this experiment can be used to inform decision-making and strategic planning. By understanding the key themes and concepts in BSI documents related to BIM, stakeholders can align their actions and initiatives with the identified focus areas. This alignment enhances performance by ensuring that efforts are targeted at the most critical aspects of BIM pervasive major project delivery, such as procurement, cost optimization, and spatial coordination.

Experiment 5: Thus experiment analysed BIM policy documents, providing guidance on policy coverage and focus areas. Stakeholders can use this information to assess the relevance of existing policies to their projects and identify potential gaps or areas for improvement. By aligning project practices with established policies, stakeholders can enhance performance through compliance and adherence to recognized standards and guidelines.

Experiment 6: This experiment analysed GCS documents, shedding light on the emphasis placed on construction, cost, procurement, industry, client, and strategy. These findings can guide stakeholders in developing effective project strategies and approaches. By incorporating the identified topics, such as optimizing procurement, cost, and digital capability, stakeholders can enhance performance by aligning their actions with the strategic goals outlined in GCS documents.

In summary, by leveraging the knowledge gained from these findings, project stakeholders can make informed decisions, prioritize their efforts, enhance collaboration, align with established policies, and develop effective strategies. These actions contribute to optimizing performance, enhancing efficiency, and achieving successful outcomes in BIM pervasive major project delivery.

6.3 Comparison with Existing Literature

6.3.1 Overview

The findings from the experimentation generally align with and complement the findings from the systematic literature review. The systematic literature review provided a theoretical understanding of the key concepts and themes related to BIM pervasive major project delivery, while the experimentation provided empirical evidence and insights derived from real-world data.

The systematic literature review identified the importance of team performance, learning legacies, technical solutions, industry guides and standards, potentials of BIM execution, applied AI, and the theory of DIK in the context of BIM pervasive major project delivery. The experimentation findings corroborated these theoretical insights and further enhanced the understanding of these topics in practical applications.

For instance, the systematic literature review highlighted the potential of BIM in enhancing team performance and improving project outcomes (Assaf and Hassannain, 2014). The experimentation findings confirmed these claims by showcasing the benefits of BIM in terms of productivity improvement, collaboration enhancement, and time and cost savings. The experiments also provided specific technical solutions enabled by BIM, such as clash detection, visualization, and quantity estimation, which contribute to performance enhancement in major construction project delivery.

Similarly, the systematic literature review emphasized the significance of industry guides and standards, applied AI, and DIK theory in BIM execution (Blanco et al., 2020). The experimentation findings supported these claims by demonstrating the importance of standardized protocols, AI applications for decision-making, and the integration of data, information, and knowledge in BIM pervasive major project delivery.

6.3.2 Addressing Identified Research Gaps

The experimental studies effectively addressed some of the gaps identified in the systematic literature review, thereby enhancing the understanding of BIM pervasive major project delivery, as described below.

Experiment 1: This experiment addressed the gap related to the geographic distribution of BIM research and publications. The systematic literature review highlighted the limited presence of African countries in BIM research, indicating challenges in BIM adoption. The experimental study confirmed this finding by identifying regions such as Asia, the USA, Europe, New Zealand, and Australia as the main contributors to BIM publications. By shedding light on this gap, the experiment provided empirical evidence that can inform efforts to bridge the research gap in African countries and facilitate knowledge sharing across regions.

Experiment 2: This experiment focused on the use of web scraping and TM techniques to study the use of BIM in enhancing performance in major construction project delivery. This approach directly addressed the gap in the systematic literature review regarding the need for data-driven insights into BIM's impact on performance. By analysing BIM-related text data and identifying frequent words related to performance, efficiency, time, quality, and cost, the experiment provided empirical evidence of BIM's influence on these performance indicators. The visualization of topics further elucidated the specific areas within major construction project delivery where BIM can contribute to performance enhancement. Thus, the experiment effectively filled the gap by providing data-driven insights into the relationship between BIM and performance.

Experiment 3: This experiment leveraged TM techniques to analyse frequently used words and their relationships in the corpus. This addressed the gap in the systematic literature review related to understanding the associations between different concepts in the context of BIM pervasive major project delivery. By uncovering these associations and visualizing the relationships, the experiment provided valuable insights into the interconnectedness of various project elements. This knowledge can help stakeholders improve performance by enhancing communication, collaboration, and coordination among different aspects of major construction project delivery.

Experiment 4 and **Experiment 5:** The two experiments focused on analysing BSI documents and BIM policy documents, respectively. These experiments addressed the gap in the systematic literature review regarding the coverage and focus areas of industry standards and government policies related to BIM. By applying TM techniques to these documents, the experiments identified key themes, concepts, and priorities. This empirical analysis provided concrete evidence of the content and emphasis within these documents, allowing stakeholders to align their practices and strategies with established standards and policies. This filling of the gap facilitates the effective utilization of industry guides and standards and ensures compliance with governmental directives.

In summary, the experimental studies effectively leveraged the identified gaps in the systematic literature review by providing empirical evidence, data-driven insights, and specific analysis of BIM-related text data, industry documents, and government policies. By addressing these gaps, the experiments enhanced the understanding of BIM pervasive major project delivery and provided valuable information that can be utilized to bridge research gaps, optimize performance, and facilitate successful implementation in this domain.

6.4 Integration of BIMTAPE Framework and TM Experiments with Key Findings and Recommendations

The integration of the BIMTAPE Framework's TM experiments with the key findings and recommendations of the Transportation Capital Projects Delivery Study (Waldman, 2023) offers a potent amalgamation to bolster major project delivery. By incorporating TM techniques, the framework adds a dimension of data-driven insights that can enhance decision-making, communication, and risk management. The research conducted by (Bareka et al., 2019, 2021, and 2022; Zaed et al., 2022a and 2022b) further fortifies this synthesis.

6.4.1 Alignment with Key Findings and Recommendations

6.4.1.1 Key Finding 1: Streamlining Decision-Making and Communication

The BIMTAPE Framework's integration of TM techniques directly complements the study's focus on streamlining decision-making and communication. TM's ability to extract trends from BIM research publications and documents offers stakeholders timely insights for informed decision-making. By leveraging TM analyses (Bareka et al., 2021 and 2022), organizations can align with the study's objective of fostering the "One City" project delivery approach through enhanced communication and shared knowledge.

6.4.1.2 Key Finding 2: Consistent Project Management Procedures

The TM experiments within the BIMTAPE Framework echo the study's emphasis on consistent project management procedures. By mining and analyzing text data, the framework aids in identifying recurring terms and themes, facilitating the establishment of standardized project management practices. This integration (Bareka et al., 2019 and 2021) ensures alignment with the study's recommendation for uniform project management protocols.

6.4.1.3 Key Finding 3: Enhanced Training and Access to Tools

The BIMTAPE Framework's incorporation of TM techniques resonates with the study's call for enhanced training and tool access. By employing TM insights, project managers gain a comprehensive understanding of performance-related terms, such as efficiency, cost, and quality. This knowledge aligns with the study's recommendation to empower project managers (Bareka et al., 2022; Zaed et al., 2022b) with tools and resources for effective project tracking and delivery.

6.4.1.4 Key Finding 4: Efficient Hiring and Procurement

The BIMTAPE Framework's utilization of TM analyses aligns with the study's recommendation for efficient hiring and procurement. By mining industry documents, the framework assists in identifying industry trends, helping organizations expedite the hiring of project staff and streamline consultant procurement. The integration (Zaed et al., 2021) underscores the study's objective of optimizing resource allocation.

6.4.1.5 Key Finding 5: Collaborative Risk Management

The BIMTAPE Framework's incorporation of TM techniques complements the study's call for collaborative risk management. By analyzing text data, the framework identifies risk-related terms and trends, aiding in risk assessment and mitigation. This integration enhances the study's recommendation for interdepartmental risk analysis (Bareka et al., 2022), contributing to more effective risk management.

6.4.2 Integration with Key Recommendations

6.4.2.1 Establishment of a Capital Project Management Office (CPMO)

The BIMTAPE Framework's integration of TM techniques aligns with the study's proposal for a CPMO. By analyzing industry documents and research trends, the framework informs the decisions of the CPMO, providing valuable insights (Bareka et al., 2021) for effective project oversight and decision-making.

6.4.2.2 Strengthening Construction Cost Estimating Processes

The BIMTAPE Framework's utilization of TM techniques resonates with the study's recommendation to strengthen construction cost estimating processes. Through TM analyses of construction-related terms (Bareka et al., 2019 and 2021), the framework assists in detailed cost reviews, ensuring that the cost estimating processes are precise and comprehensive.

6.4.2.3 Enhanced Right of Way and Utility Investigation Programs

The BIMTAPE Framework's integration of TM techniques aligns with the study's emphasis on enhanced right of way and utility investigation programs. By extracting insights from relevant documents, the framework contributes to utility identification services, aiding in comprehensive project designs and risk assessment.

6.4.2.4 Collaborative Interdepartmental Risk Review and Management

The BIMTAPE Framework's incorporation of TM techniques complements the study's call for collaborative interdepartmental risk review and management. By analyzing risk-related terms and trends (Bareka et al., 2022), the framework equips organizations with data-driven insights, facilitating cross-disciplinary collaboration and risk management.

6.5 Implications and Future Industrial and Academic Directions

The alignment between the experimentation findings and the systematic literature review yields significant implications for both the construction industry and future research endeavours. The consistent alignment between theoretical insights and empirical evidence fortifies the foundation of knowledge, enhancing the credibility and practical relevance of the research outcomes.

The identified implications underscore the potential of BIM to revolutionise major construction project delivery. By fostering improved team collaboration, optimizing project cost and schedule, and harnessing the capabilities of AI technologies, BIM emerges as a potent tool for elevating project performance. The industry stands to gain substantial benefits by embracing BIM practices, adhering to industry standards and guides, and embracing technological advancements. This comprehensive approach has the potential to unlock the full capabilities of BIM, ultimately contributing to the attainment of successful project outcomes.

Furthermore, future research directions should delve into specific challenges and barriers surrounding BIM adoption. Investigating the effectiveness of various BIM implementation strategies and assessing the long-term impacts of BIM on project performance and sustainability would provide invaluable insights. Moreover, exploring the integration of BIM with emerging technologies such as the IoT, blockchain, and virtual reality promises to shape the future landscape of BIM pervasive major project delivery.

The harmonious integration of the BIMTAPE Framework's TM experiments with the key findings and recommendations of the Transportation Capital Projects Delivery Study bolsters the framework's effectiveness. By infusing data-driven insights into decision-making, communication, and risk management, the BIMTAPE Framework presents a holistic approach to enhancing major project delivery. The confluence of empirical evidence from TM experiments with the study's recommendations offers a robust strategy for achieving project delivery excellence.

As the trajectory moves forward, the continuous exploration of how TM techniques can further enhance the BIMTAPE Framework's applicability across diverse industries remains a promising avenue of research. This endeavour holds the potential to usher in a new era of project delivery marked by informed decision-making, collaborative communication, and proactive risk management. Through this ongoing exploration, the framework's capacity to drive project success stands poised for continuous refinement and expansion.

6.6 Summary

This chapter presented the findings of the experimentation conducted in this study and compared them with the findings from the systematic literature review. The alignment and consistency between the empirical results and theoretical insights demonstrate the potential of BIM in enhancing performance in major construction project delivery. The implications of the findings highlight the importance of BIM adoption, industry guides and standards, AI applications, and the integration of data, information, and knowledge for successful BIM pervasive major project delivery. Future research should further explore the challenges, implementation strategies, and long-term impacts of BIM in the construction industry.

Chapter 7
Conclusion

7.1 Introduction

The adoption of BIM in the construction industry has been recognized for its potential to enhance performance in major construction project delivery. This chapter provides a comprehensive conclusion by summarizing the key findings and contributions of the research. The objectives of the study are addressed individually, highlighting the outcomes and their significance in the context of BIM pervasive major project delivery.

7.2 Remarks on Research Objectives

7.2.1 Research Objective 1

The systematic literature review conducted in this research provided a solid foundation for understanding the key concepts and themes related to BIM pervasive major project delivery. It justified the use of TM as a valuable approach to analyse BIM-related texts and extract insights for performance enhancement. The review confirmed the global adoption of BIM and identified geographical variations in research trends. It also highlighted the benefits, challenges, and limitations associated with BIM implementation. By synthesizing existing research, the systematic literature review justified the need for TM and its potential to contribute to improved project outcomes.

7.2.2 Research Objective 2

To achieve this objective, a TM methodology and the BIMTAPE Framework were developed. The methodology involved techniques such as web scraping, term-document matrices, clustering, and topic modelling using LDA. These techniques enabled the extraction of valuable insights from BIM-related texts and the identification of key themes and concepts. The BIMTAPE Framework provided a structured approach for leveraging text analysis to enhance project management in the context of BIM pervasive major construction projects. It emphasized the importance of model deployment, documentation, communication, and model updates in utilizing BIM-related texts for performance enhancement.

7.2.3 Research Objective 3

The experimental studies conducted in this research demonstrated the effectiveness of the developed TM methodology for major construction project delivery. Experiment 1 examined the geographic distribution of BIM research and publications, providing empirical evidence of research trends and identifying regions with limited BIM adoption. Experiment 2 focused on web scraping and TM techniques to study the use of BIM in enhancing performance. The findings highlighted the relevance of BIM in different stages of construction and its impact on performance indicators. Experiment 3 analysed frequently used words and their relationships, providing insights into project elements and their interconnections. These experiments validated the effectiveness of the TM methodology in extracting valuable insights from BIM-related texts and contributing to performance improvement in major construction projects.

7.3 Overview of the Work

The conducted study has revealed that the use of BIM pervasive major project delivery is associated with benefits and limitations. Several BIM impacts are realized in the construction sector, which need to be managed by the construction managers. AEC companies in the USA, UK, France, and Germany have positively indicated that BIM increases productivity and improves the quality of works. The technology leads to better quality, increased coordination, saving on cost and time, and increased collaboration. These revelations are provided through the approach of TM, which is depicted to be an effective strategy of indicating how BIM can be successful in performance enhancement of project delivery. The project also reveals that BIM executions are based on industry standards and guides. These set the protocols that have to be followed and utilized in the prescription of how BIM has to be implemented when providing service to users. The protocols also facilitate in ensuring that construction contracts have the capability and potential to realize the desired value for all stakeholders. The standards and guides are further used by government agencies to ensure that regulatory frameworks are embraced in the use of BIM.

The project reveals that BIM adoption needs to have new methodologies, tools, and processes that support big project teams. The approach is undertaken to ascertain that there is significant improvement in the effectiveness of the project. BIMTAPE

Framework is useful in the analysis of different BIM-related texts that include technical documents, standards, and specifications. These have to be accomplished by adhering to the stages of model deployment, documentation, communication, and model updates. These stages facilitate in ascertaining that BIM text analysis is accomplished effectively to determine its contributory role in supporting performance in major pervasive project delivery.

7.4 Concluding Insights

In conclusion, this research has demonstrated the effectiveness of TM techniques and the BIMTAPE Framework in enhancing performance in BIM pervasive major project delivery. The integration of empirical evidence from experimental studies with theoretical insights from the systematic literature review has provided a comprehensive understanding of BIM and its potential in improving project outcomes. The developed methodology and framework offer practical approaches to leverage BIM-related texts and extract valuable insights for informed decision-making and enhanced project management. The contributions of this research have implications for practitioners, researchers, and policymakers, and further research directions have been identified to continue advancing the field of BIM pervasive major project delivery.

Additionally, SNA was employed in this research to create network analyses of the 20 frequent words generated from each experiment, with the exception of Experiment 2, which utilized a real-time case scenario. SNA provided quantitative data that allowed the researcher to examine the relationships and interactions among the identified words and concepts. This analysis further enriched the understanding of the interconnectivity and importance of these words within the context of BIM pervasive major project delivery.

The utilization of SNA offered valuable insights into the network structures and patterns formed by the frequent words, highlighting the key associations and dependencies among them. By visualizing these networks, the researcher was able to identify central nodes, clusters, and influential factors that play a significant role in BIM pervasive major project delivery. The quantitative data obtained from SNA provided a deeper understanding of the relationships between various concepts, enabling better informed decision-making and strategic planning.

The integration of SNA with TM, the BIMTAPE Framework, and other methodologies used in this research has provided a comprehensive and multifaceted approach to studying BIM pervasive major project delivery. The combination of qualitative and quantitative analyses has enriched the research findings, offering a holistic understanding of the complexities and dynamics within the construction industry. These insights have important implications for stakeholders across various domains.

Policy makers and government bodies can use these findings to develop and refine regulations, guidelines, and standards that promote the adoption and implementation of BIM in major construction project delivery. The construction industry as a whole can benefit from the enhanced understanding of BIM's potential in improving project outcomes, optimizing processes, and fostering collaboration among stakeholders.

Companies and professionals in the construction sector can leverage the research findings to inform their decision-making processes, enhance project management strategies, and optimize resource allocation. By embracing TM, the BIMTAPE Framework, and SNA, organizations can gain a competitive advantage by extracting valuable insights from textual data, identifying KPIs, and optimizing project delivery.

Furthermore, the public and end-users of construction projects stand to benefit from the implementation of BIM pervasive major project delivery approaches informed by this research. Improved project outcomes, such as increased efficiency, reduced costs, enhanced quality, and better communication, contribute to the creation of sustainable and user-centric built environments.

In conclusion, the integration of TM, the BIMTAPE Framework, and SNA has provided a robust and comprehensive approach to studying BIM pervasive major project delivery. The research findings and insights generated through these methodologies offer practical implications for stakeholders, including policy makers, the construction industry, companies, professionals, and the public. By leveraging these findings, stakeholders can enhance project outcomes, optimize processes, and contribute to the advancement of the construction industry. The identified research gaps and further research directions highlight the need for continued exploration and advancement in the field of BIM pervasive major project delivery.

7.5 Contributions

The study informs both professional practice and academic research on issues of text analysis as a technical solution to challenges of error, uncertainty, and expected in BIM pervasive major project delivery. TM approach has the possibility of ensuring that BIM pervasive major project delivery performance is enhanced. The approach is effectively demonstrated through the use of BIMTAPE Framework. The study also informs that BIM plays a significant role in increasing the performance output of the construction projects. The tool is critical in supporting the improvement in the productivity of the construction projects by ascertaining that there is effective collaboration that exists among the different stakeholders. The conducted study also indicates that advanced solutions need to be embraced in the major construction projects and delivery. These solutions facilitate in ascertaining that project delays and cost overruns are addressed and there are no uncertainties and complexities linked to the construction projects. Furthermore, recent trends in globalization support digitalization of all systems, which indicates that construction industry has to align itself to the digital economy by embracing the technology of BIM. Such has to occur along the technologies of AI and ML.

The construction practice and policy need to incorporate BIM technology as a means of enhancing efficiency in service delivery to the customers. The industry needs to embrace a fast change to ensure that it becomes more environmentally friendly, sustainable, and efficient. Government agencies across the globe should embrace BIM technology and standards to ensure that the tool becomes a means of enhancing the attainment of the desired success level. The approach will also contribute to the increase in the level of efficiency and sustainability in the implementation of realistic and practical policies. Furthermore, construction firms across the globe will stand a chance to enjoy the benefits attributed to the use of BIM technology.

7.6 Research Limitations

It is important to be transparent about limitations, as they can help to contextualize the research outcomes for caution among practitioners, and indicate areas where further research is needed. One limitation of this research is that it focuses on a specific TM approach to BIM pervasive major project delivery, and may not be generalizable to other approaches or methodologies. Additionally, the research may not be able to capture all of the complexities and nuances of BIM adoption in construction projects, which could limit the generalizability of the findings.

Another limitation is that the research is based on a limited number of case studies, which may not be representative of other projects or industries. This could limit the generalizability of the research outcomes. The research may be limited by the availability and quality of data, which could impact the accuracy of the findings. Finally, the research may not be able to fully capture the impact of human factors such as team dynamics, communication and other subjective elements which play an important role in the implementation of BIM.

7.7 Future Research Directions

BIM has been depicted to have limited adoption and implementation across the globe, which indicates that the technology is not commonly and widely utilized in the construction sectors. Future research should evaluate how BIM use is affected by the recent technological advances and revolutions that are recorded worldwide, and seek to establish the resources that are needed to ensure that BIM has the capability of realizing the anticipated success in the market environment. This includes the technological integrations that have to be embraced in major pervasive projects to ascertain that BIM has the potential to realize optimal performance.

The current experiments have identified a list of prioritized keywords, which could be explored in more long-term studies. Furthermore, it is recommended to use these keywords to both qualitatively and quantitatively interpret the status of BIM oriented research and BIM pervasive practice. This quantitative interpretation can be referred to as the “BIM Momentum Index”, and can be calculated by aggregating the frequency of the top 10 or 20 keywords found in academic papers and professional documents within the six main types of documentation. Sub-indexes, such as “BIM Research Momentum Index”, “BIM Practice Momentum Index”, and “BIM Strategy Momentum Index” could be established to analyse the achievements in BIM research, professional practice, and governance at the national and/or enterprise level. This calculation will be based on annual data of the selected keywords, allowing us to observe the trends in BIM advancements over time.

Based on the research findings, it is recommended that stakeholders, including policy makers, should consider the potential benefits of adopting BIM and TM approaches in major construction projects. These benefits include improved decision-making, cost savings, increased efficiency, and improved project outcomes, which can ultimately lead to a more sustainable and successful construction industry.

Furthermore, the establishment of a “BIM Momentum Index” as suggested can provide valuable insights for policy makers and industry professionals to monitor the progress of BIM adoption and identify areas for improvement. This index can also help to bridge the gap between academic research and practical implementation, allowing for a more holistic approach to BIM integration in the construction industry. In addition, companies and professionals in the construction industry can benefit from the use of BIM and TM approaches by enhancing their project management and communication processes, as well as improving their ability to meet project deadlines and budgets.

Finally, the public can benefit from the use of BIM and TM approaches through improved building safety, reduced environmental impact, and increased transparency in the construction process.

7.8 Recommendations for Stakeholders

7.8.1 Policy Makers and Government Bodies

Policy makers and government bodies play a critical role in shaping the construction industry and facilitating the widespread adoption of BIM. Based on the research findings, the following recommendations are proposed.

7.8.1.1 Develop Comprehensive BIM Policies and Guidelines

Policy makers should collaborate with industry experts and stakeholders to establish clear and comprehensive BIM policies and guidelines. These should address key aspects such as BIM standards, data interoperability, information exchange protocols, and project requirements. By providing a structured framework, policy makers can ensure consistency, efficiency, and interoperability across BIM projects.

7.8.1.2 Foster Industry Collaboration

Encourage collaboration among government agencies, industry associations, and private sector stakeholders to drive BIM implementation, and facilitate platforms for knowledge sharing, best practice exchange, and capacity building. By fostering collaboration, policy makers can facilitate the dissemination of BIM expertise and support the development of a skilled workforce.

7.8.1.3 Provide Financial Incentives and Support

Introducing financial incentives, grants, and funding programs can encourage companies and professionals to invest in BIM adoption. Offering support for training programs, research and development initiatives, and pilot projects promotes innovation and knowledge transfer. By providing financial support, policy makers can lower the barriers to BIM implementation and accelerate its adoption.

7.8.2 Industry as a Whole

The construction industry as a whole can benefit from the findings of this research. The following specific recommendations are proposed.

7.8.2.1 Embrace a Collaborative and Integrated Approach

Encourage stakeholders, including architects, engineers, contractors, and suppliers, to adopt a collaborative and integrated approach to project delivery. Foster a culture of information sharing, open communication, and collaboration to maximize the benefits of BIM. This can result in improved coordination, reduced rework, and enhanced project outcomes.

7.8.2.2 Invest in BIM Training and Education

Promote continuous professional development and training programs to equip industry professionals with the necessary skills and knowledge in BIM. Encourage universities and educational institutions to incorporate BIM into their curricula to produce a new generation of skilled BIM practitioners. By investing in training and education, the industry can build a capable workforce that is proficient in BIM technologies and processes.

7.8.2.3 Promote Data Standardization and Interoperability

Advocate for the adoption of open standards and interoperable platforms to enable seamless data exchange and integration. Encourage the use of common data environments (CDEs) and cloud-based platforms to facilitate collaboration and data sharing among project stakeholders. By promoting data standardization and interoperability, the industry can eliminate data silos, improve efficiency, and enable more effective decision-making.

7.8.3 Companies

Companies operating in the construction industry can leverage the findings of this research to enhance their BIM implementation strategies. The following recommendations are proposed.

7.8.3.1 Develop BIM Implementation Plans

Create comprehensive BIM implementation plans that align with organizational goals and project requirements. These plans should outline the necessary steps, resources, and timelines for successful BIM adoption. Companies should consider factors such as staff training, technology infrastructure, and collaboration processes when developing their implementation plans.

7.8.3.2 Invest in BIM Technology and Infrastructure

Allocate resources to invest in BIM software, hardware, and infrastructure that support the effective implementation of BIM. Ensure that the technology and tools selected are compatible with industry standards and enable seamless collaboration and data exchange. Companies should also consider cloud-based solutions to enhance accessibility and scalability.

7.8.3.3 Foster a BIM Culture and Mindset

Cultivate a culture that embraces BIM and encourages employees to adopt BIM methodologies and processes. Promote continuous learning, knowledge sharing, and innovation within the organization. Develop internal BIM champions who can drive the adoption of BIM and act as ambassadors for change.

7.8.4 Professionals

Professionals working in the construction industry can leverage the research findings to enhance their individual BIM competencies and contribute to the overall success of BIM projects. The following recommendations are proposed.

7.8.4.1 Continuously Update BIM Knowledge and Skills

Stay updated with the latest trends, technologies, and best practices in BIM. Engage in professional development activities, attend industry conferences, and participate in training programs to enhance BIM competencies. Embrace lifelong learning to remain at the forefront of BIM advancements.

7.8.4.2 Collaborate and Share Knowledge

Engage in collaborative platforms and networks to exchange knowledge, experiences, and lessons learned with peers in the industry. Participate in industry associations, online forums, and communities to share insights and contribute to the collective growth of the BIM community.

7.8.4.3 Embrace Interdisciplinary Collaboration

Recognize the importance of interdisciplinary collaboration and actively engage with other professionals involved in BIM projects (particularly those in other AEC professions, but also other major project stakeholders). Foster effective communication, teamwork, and coordination to achieve project goals and deliver value to clients. Embrace a holistic approach that goes beyond individual professional boundaries.

7.8.5 The Public

The public can benefit from the implementation of BIM in major construction project delivery. The following recommendations are proposed.

7.8.5.1 Foster Transparency and Accountability

Encourage project owners and public institutions to prioritize transparency and accountability in major projects that utilize BIM. Ensure that project information, progress reports, and outcomes are accessible to the public. This fosters trust and allows for greater scrutiny of project performance.

7.8.5.2 Promote Sustainability and Environmental Considerations

Advocate for the integration of sustainable design principles and environmental considerations in BIM projects. Encourage the use of energy-efficient materials, renewable energy sources, and sustainable construction practices. BIM can facilitate the simulation and analysis of sustainability measures, enabling informed decisions that benefit the environment.

7.8.5.3 Enhance Public Awareness and Education

Educate the public about the benefits and potential of BIM in major construction project delivery. Increase awareness of how BIM can lead to better project outcomes, reduced costs, improved safety, and minimized environmental impact. By enhancing public awareness, stakeholders can garner support and create a favourable environment for BIM implementation.

By implementing these recommendations, stakeholders across various sectors can contribute to the successful adoption and application of BIM in major construction project delivery. These recommendations foster collaboration, innovation, and sustainability, ultimately leading to improved project outcomes, increased efficiency, and enhanced value for all stakeholders involved.

References

Aaltonen, K. (2011). Project stakeholder analysis as an environmental interpretation process. *International Journal of Project Management*, 165–183. https://doi.org/10.1016/j.ijproman.2010.02.001.

Aaltonen, K., Kujala, J., Lehtonen, P., and Ruuska, I. (2010). A stakeholder network perspective on unexpected events and their management in international projects. *International Journal of Managing Projects in Business*, *3*, 564–588. https://doi.org/10.1108/17538371011076055

Abbasnejad, B., and Moud, H. I. (2013). BIM and basic challenges associated with its definitions, interpretations and expectations. *International Journal of Engineering Research and Applications*, (2), 287–294.

Abu-Qulbain, M., and Hill, S. (2016). A review of reactive electrical power and power factor correction (PFC) in the Crossrail low-voltage (LV) systems. *Crossrail Ltd*. Available at: https://learninglegacy.crossrail.co.uk/documents/a-review-of-reactive-electrical-power-and-power-factor-correction-pfc-in-the-crossrail-low-voltage-lv-systems/.

Abu-Qulbain, M., and Hill, S. (2021). A Review of Reactive Electrical Power and Power Factor Correction (PFC) in the Crossrail Low Voltage (LV) Systems. *Crossrail Ltd*, ICE Publishing. Available at: https://learninglegacy.crossrail.co.uk/documents/a-review-of-reactive-electrical-power-and-power-factor-correction-pfc-in-the-crossrail-low-voltage-lv-systems/.

Ademci, E., and Gundes, S. (2018). Review of studies on BIM adoption in AEC industry. In E. Ademci and S. Gundes (Eds.), *Review of studies on BIM adoption in AEC industry, 5th International Project and Construction Management Conference (IPCMC) Proceedings* (pp. 1046–1055). Istanbul Technical University.

Adrian, J. J. (1995). *Construction productivity: Measurement and improvement*. Stipes.

Aggarwal, C. C., and Zhai, C. (Eds.). (2012). *Mining text data*. Boston, MA: Springer US. https://doi.org/10.1007/978–1-4614–3223–4

Agrawal, R., Mannila, H., Srikant, R., Toivonen, H., and Verkamo, A. I. (1996). Fast discovery of association rules. *Advances in Knowledge Discovery and Data Mining*, *12*(1), 307–328.

Ahmad, T. (2018). Corporate Information Handover Project: Enabling Data Migration. *Crossrail Ltd*. Available at: https://learninglegacy.crossrail.co.uk/documents/corporate-information-handover-project-enabling-data-migration/.

Ahmed, S., Thaheem, M.J., and Shrestha, A. (2018). Automatic defect detection and quantification in construction sites using convolutional neural networks. *Automation in Construction*, 89, pp.24-37.

Ahn, Y. H., Kwak, Y. H., and Suk, S. J. (2016). Contractors' transformation strategies for adopting building information modelling. *Journal of Management in Engineering*, *32*(1), Article 05015005. https://doi.org/10.1061/(ASCE)ME.1943–5479.0000390

Aho, A. V. (1991). Algorithms for finding patterns in strings. In M. Y. Vardi (Ed.), *Handbook of theoretical computer science, vol. A: Algorithms and complexity* (pp. 255–300). Cambridge, MA: MIT Press.

Ahuja, R., Sawhney, A., and Arif, M. (2017). Prioritizing BIM capabilities of an organization: An interpretive structural modelling analysis. *Procedia Engineering*, *196*, 2–10.

AIA California Council. (2007). *Integrated project delivery: A working definition*. http://aiacc.org/wp-content/uploads/2010/07/A-Working-Definition-V2-final.pdf

Al Qady, M., and Kandil, A. (2014). Automatic clustering of construction project documents based on textual similarity. *Automation in Construction*, *42*, 36–49. https://doi.org/10.1016/j.autcon.2014.02.006

Alaka, H. A., Oyedele, L. O., Owolabi, H. A., Oyedele, A. A., Akinade, O. O., Bilal, M., and Ajayi, S. O. (2017). Critical factors for insolvency prediction: Towards a theoretical model for the construction industry. *International Journal of Construction Management*, *17*, 25–49. https://doi.org/10.1080/15623599.2016.1166546

Alharahsheh, H. and Pius, A. (2020) 'A Review of Key Paradigms: Positivism vs Interpretivism', *Journal of Contemporary Issues in Business and Government*, 26(2), pp. 76-84.

Al-Hussein, M. (2019). 36th International Symposium on Automation and Robotics in Construction and Mining (ISARC, 2019).

Ali Musarat, S., Jamil, M., Tufail, R., Ahmad, R., and Rahman, S. (2022). Digitalisation in the construction sector: A systematic review. *Journal of Construction Engineering and Management*, 148(1), pp.04021109.

Ali, K., Alhajlah, H., and Kassem, M. (2022). Collaboration and risk in building information application in construction. *Journal of Civil Engineering and Management*, *27*(5), 303–315.

Alkhiami, M. I. B., and Zangana, A. (2017). *The impact of BIM on design phase productivity: The West Link case study* [Master's thesis, Chalmers University of Technology]. Chalmers Open Digital Repository. https://odr.chalmers.se/server/api/core/bitstreams/b0cc5a57-4596-419c-b4be-a72827aaff46/content

Allmon, E., Haas, C. T., Borcherding, J. D., and Goodrum, P. M. (2000). U.S. construction labor productivity trends, 1970–1998. *Journal of Construction Engineering and Management*, *126*, 97–104. https://doi.org/10.1061/(ASCE)0733–9364(2000)126:2(97)

Alloghani, E., Al-Dubai, A. and Al-Fuqaha, A., 2020. A survey of classification techniques for text data. *Journal of Computer Science and Technology*, 35(3), pp.311-323.

Alsubaey, M., Asadi, A., and Makatsoris, H. (2015). A Naïve Bayes approach for EWS detection by text mining of unstructured data: A construction project case, In *2015 SAI Intelligent Systems Conference (IntelliSys)* (pp. 164–168). New York, NY: IEEE. https://doi.org/10.1109/IntelliSys.2015.7361140

Alsulaimi, A., and Abdullah, T. (2020). Management of stakeholder communications in IT projects. In *2020 3rd International Conference on Computer Applications and Information Security* (ICCAIS) (pp. 1–6). IEEE. http://dx.doi.org/10.1109/ICCAIS48893.2020.9096842

Amatriain, X., and Basilico, J. (2012). *Netflix recommendations: Beyond the 5 stars.* In Proceedings of the 5th ACM Conference on Recommender Systems (pp. 4-11). ACM. https://doi.org/10.1145/2365952.2365955

Amazon Web Services. (2024). What is OCR (Optical Character Recognition)?. Available at: https://aws.amazon.com/what-is/ocr/ [Accessed 3 April 2024]

Anderson, M., Leary, N., and King, J. (2018). Earthing and Bonding. *Crossrail Ltd*, ICE Publishing. Available at: https://learninglegacy.crossrail.co.uk/documents/earthing-and-bonding/.

Andriamamonjy, A., Saelens, D., Klein, R. (2019). A combined scientometric and conventional literature review to grasp the entire BIM knowledge and its integration with energy simulation. *Journal of Building Engineering*, *22*, 513–527.

Angelos, S., Ryburn, J., and Roberts, J. (2016). 3D geological model of the completed Farringdon underground railway station. *Crossrail Ltd*. Available at: https://learninglegacy.crossrail.co.uk/documents/3d-geological-model-completed-farringdon-underground-railway-station/.

Arayici, Y., Egbu, C. O., and Coates, S. P. (2012). Building information modelling (BIM) implementation and remote construction projects: Issues, challenges, and opportunities. *Architectural Engineering and Design Management*, *8*(2), 120–129.

Aria, M., and Cuccurullo, C. (2017). A brief introduction to bibliometrix. *Journal of Informetrics*, *11*(4), 959–975.

Ashcraft, H. W. Jr. (2011). *Negotiating an integrated project delivery agreement*. San Francisco, CA: Hanson Bridgett LLP. https://images.autodesk.com/latin_am_main/files/negotiating_an_ipd_agreement.pdf

ASQ. (2022). *What is the nine windows technique?*. https://asq.org/quality-resources/nine-windows

Assaf, S., and Hassanain, M. (2014). Effectiveness of project teams and their impact on the application in construction. *Journal of Civil Engineering and Management*, *27*(5), 303–315.

Asvadurov, S., Brinded, T., and Ellis, M. (2018). Transforming infrastructure performance: An owner's perspective. *Global Infrastructure Initiative*. https://www.globalinfrastructureinitiative.com/sites/default/files/voices/edition/pdf/June-2018-Voices-0709.pdf

Awakul, P., and Ogunlana, S. O. (2002). The effect of attitudinal differences on interface conflict on large construction projects. *Environmental Impact Assessment Review*, *22*, 311–335. https://doi.org/10.1016/S0195–9255(02)00007–0

Azhar, S. (2011). Building information modelling (BIM): Trends, benefits, risks, and challenges for the AEC industry. *Leadership and Management in Engineering*, *11*, 241–252. https://doi.org/10.1061/(ASCE)LM.1943–5630.0000127

Azhar, S., Brown, J., and Farooqui, R. (2009). BIM-based sustainability analysis: An evaluation of building performance analysis software. *Proceedings of the 45th ASC Annual Conference*. New York: ACM.

Baccarini, D. (1999). The logical framework method for defining project success. *Project Management Journal*, *30*(4), 25–32.

Badran, D., AlZubaidi, R., and Venkatachalam, S. (2020). BIM based risk management for design bid build (DBB) design process in the United Arab Emirates: A conceptual framework. *International Journal of Systems Assurance Engineering and Management*, *11*, 1339–1361. https://doi.org/10.1007/s13198–020–00993-w

Baek, S., Wang, Q., and AbouRizk, S. (2021). Knowledge-based text mining approaches to support text-based information extraction for construction and project management. *Automation in Construction*, 124, p.103567.

Baiden, B., and Price, A. (2011). The effect of integration on project delivery team effectiveness. *International Journal of Project Management*, *29*(2), 129–136.

Bailey, R., Steele, M., and Braddish, P. (2016). Covered Way 126: Safeguarding a Brittle Structure over a Live Railway during Ground Movements. *Crossrail Ltd*. Available at: https://learninglegacy.crossrail.co.uk/documents/covered-way-126-safeguarding-brittle-structure-live-railway-ground-movements/.

Bajpai, N. (2011) *Business Research Methods*. New Delhi: Pearson Education India.

Bannister, P. (2004). Designing buildings that actually perform. *Environment Design Guide*, *18*.

Bareka, A. (2022). *Unlocking Insights: A Curated Dataset and Text Mining Analysis of BIM in Major Construction Project Delivery*. University of Strathclyde, V1, https://doi.org/10.17632/yk7vgxtjgp.1.

Bareka, A. (2023), *Web scraping*. University of Strathclyde, V3, https://doi: 10.17632/5hcg7tddc4.3.

Bareka, A., and Chen, Z. (2019). A lifecycle-oriented SNA framework for BIM adoption in major construction project. *Proceedings of the Doctoral Workshop on Contemporary Advances in Research Methodology in Construction Management* (ARCOM). 01/11/2019, Glasgow, UK.

Bareka, A., Chen, Z., Agapiou, A., and Dimitrijevic, B. (2021). A lifecycle social network analysis framework for BIM adoption in major construction projects. *Proceedings of the 5th Coventry University Construction Conference.* 17/03/2021, Coventry, UK.

Bareka, A., Chen, Z., Agapiou, A., and Dimitrijevic, B. (2021). A text analysis roadmap for BIM pervasive major project delivery. *Frontiers in Built Environment*. Manuscript under internal review.

Bareka, A., Chen, Z., Agapiou, A., and Dimitrijevic, B. (2021). A text mining approach to trend analysis on BIM research in the Middle East and North Africa (MENA). *Proceedings of the 5th Coventry University Construction Conference.* 17/03/2021, Coventry, UK.

Bareka, A., Chen, Z., Agapiou, A., and Zaed, O. (2022). A text mining approach to performance enhancement in BIM pervasive major project delivery. *Proceedings of the International Postgraduate Research Conference (IPGRC).* 04-06/04/2022, University of Salford, Salford, UK.

Barima, O. (2017). Leveraging the blockchain technology to improve construction value delivery: The opportunities, benefits and challenges. *Construction Projects*, pp. 93–112.

Basu, S. (1991). Urban building: System constraints in Calcutta. *Economic and Political Weekly*.

Bawono, A. A., von Schumann, C. M., and Lechner, B. (2021). Study of building information modelling implementation on railway infrastructure. In E. Toledo Santos and S. Scheer (Eds.), *Proceedings of the 18th International Conference on Computing in Civil and Building Engineering, Lecture Notes in Civil Engineering* (pp. 372–382). Cham: Springer International Publishing. https://doi.org/10.1007/978–3-030–51295–8_26

Beach, T., Petri, I., Rezgui, Y., and Rana, O. (2017). Management of collaborative BIM data by federating distributed BIM models. *Journal of Computing in Civil Engineering*, *31*(4), Article 657. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000657

Becerik-Gerber, B., Jazizadeh, F., Li, N., and Calis, G. (2012). Application areas and data requirements for BIM-enabled facilities management. *Journal of Construction Engineering and Management*, *138*, 431–442. https://doi.org/10.1061/(ASCE)CO.1943–7862.0000433

Becker, D., and Ghedini, F. (2005). Managing data mining projects. *Information Management and Computer Security*, 13(4), 291-303.

Becker, K., and Ghedini, C. (2005). A documentation infrastructure for the management of data mining projects. *Information and Software Technology*, *47*(2), 95–111.

Bendale, U. M. (2018). A review on assessment of project-related factors for major construction projects. *International Journal of Research in Advanced Science and Engineering Technology*, *6*, 1623–1627. https://doi.org/10.22214/ijraset.2018.5264

Bensalah, M., Elouadi, A., and Mharzi, H. (2019). Overview: The Opportunity of BIM in railway. *Smart and Sustainable Built Environment*, *8*, 103–116. https://doi.org/10.1108/SASBE-11–2017–0060

Berners-Lee, T. (2016). *Weaving the Web: The Original Design and Ultimate Destiny of the World Wide Web by Its Inventor*. Harper San Francisco.

Bernolak, I. (1997). Effective measurement and successful elements of company productivity: The basis of competitiveness and world prosperity. *International Journal of Production Economics*, *52*, 203–213. https://doi.org/10.1016/S0925–5273(97)00026–1

Bernstein, H. M., Jones, S. A., Russo, M. A., and Laquidara-Carr, D. (2014). *The business value of BIM in Australia and New Zealand: How building information modelling is transforming the design and construction industry*. Bedford, MA: McGraw Hill Construction

BibLus, 2022. What are BIM standards?. Available at: https://biblus.accasoftware.com/en/what-are-bim-standards/

Bilal, M., Oyedele, L. O., Qadir, J., Munir, K., Ajayi, S. O., Akinade, O. O., Owolabi, H. A., Alaka, H. A., and Pasha, M. (2016). Big Data in the construction industry: A review of present status, opportunities, and future trends. *Advanced Engineering Informatics*, *30*(3), 500–521.

Binbin, W., and Xiaoyan, L. (2017). Big Data, platform economy and market competition: A strategic framework. *International Journal of Automation and Computing*, *14*, 1–12.

Bird, A., Cobbing, C., Bhalla, P., Clark, D., Guillaume, B., Howden, T., Parry, G., Sullivan, R., Summers, S. (2016) 'Noise and vibration controls for the TBM and temporary construction railway', *Crossrail Ltd*., [online] Available at: https://learninglegacy.crossrail.co.uk/documents/noise-vibration-controls-tbm-temporary-construction-railway/

Bishop, C. M. (2006). *Pattern recognition and machine learning* (Vol. 4). Cham: Springer.

Bishop, J. (2012). Learning from Mistakes. *Inside Supply Management*, *23*, 12–13.

Blake, A. (2019). Dynamics of data science skills: How Can all sectors benefit from data science talent?. *Royal Society*. https://royalsociety.org/-/media/policy/projects/dynamics-of-data-science/dynamics-of-data-science-skills- report.pdf

Blanco, F. G. B., and Chen, H. (2014). The implementation of building information modelling in the United Kingdom by the transport industry. *Procedia - Social and Behavioral Sciences*, *138*, 510–520. https://doi.org/10.1016/j.sbspro.2014.07.232.

Blanco, J., Fuchs, S., Parsons, M., and Ribeirinho, M. (2020). Artificial Intelligence: Construction technology's next frontier. *McKinsey and Company*. https://www.mckinsey.com/industries/capital-projects-and-infrastructure/our-insights/artificial-intelligence-construction-technologys-next-frontier

Blayse, A. M., and Manley, K. (2004). Key influences on construction innovation. *Construction Innovation: Information, Process, Management*, *4*, 143–154.

Blei, D.M., Ng, A.Y. and Jordan, M.I., (2003). Latent dirichlet allocation. *Journal of machine Learning research*, *3*(Jan), pp.993-1022.

Bongiorno, N., Bosurgi, G., Carbone, F., Pellegrino, O., and Sollazzo, G. (2019). Potentialities of a highway alignment optimization method in an I-BIM environment. *Periodica Polytechnica-Civil Engineering*, *63*, 352–361. https://doi.org/10.3311/PPci.12220

Borgatti, S. P., Everett, M. G., and Freeman, L. C. (2002). *Ucinet for Windows: Software for social network analysis*. Harvard, MA: Analytic Technologies.

Bouras, A., Eynard, B., Foufou, S., and Thoben, K. D. (Eds.). (2015). *Product lifecycle management in the era of Internet of Things: 12th IFIP WG 5.1 international conference, PLM 2015 Doha, Qatar, October 19-21, 2015 revised selected papers*. New York, NY: Springer New York LLC.

Bourreau, M., Sadowski, B. and Lagnier, R., (2018). Data Management and Real-time Decision Support in Construction Projects. *Journal of Construction Engineering and Management*, 144(2), p.04017125.

Brent, D., Sheldon, D., and Shephard, C. (2018). What is required for the successful delivery of infrastructure projects?. *Global Infrastructure Initiative.* https://www.globalinfrastructureinitiative.com/sites/default/files/voices/edition/pdf/June-2018-Voices-0709.pdf

Bryde, D., Broquetas, M., and Volm, J. M. (2013). The project benefits of building information modelling (BIM). *International Journal of Project Management*, *31*(7), 971–980. https://doi.org/10.1016/j.ijproman.2012.12.001

Bryman, A. and Bell, E. (2011) *Business Research Methods*. 3rd edn. Oxford: Oxford University Press.

Bryman, A. (2006) *Integrating quantitative and qualitative research: How is it done?* Qualitative Research, 6(1), pp.97-113.

Buehler, M.J., (2024). *Accelerating Scientific Discovery with Generative Knowledge Extraction, Graph-Based Representation, and Multimodal Intelligent Graph Reasoning*. Massachusetts Institute of Technology (MIT), arXiv preprint arXiv:2403.11996v3.

Bui, N. (2020). Implementation of building information modelling in Vietnamese infrastructure construction: A case study of institutional influences on a bridge project. *Electronic Journal of Information Systems in Developing Countries*, *86*(4), Article e12128. https://doi.org/10.1002/isd2.12128

BuildingSMART (2020). *Industry Foundation Classes (IFC): Enabling Open Standards for the AEC Industry*. Available at: https://www.buildingsmart.org/standards/bsi-standards/industry-foundation-classes/ (Accessed: October 2024).

Bulbrook, D., Coleman, S.-R., and Selman, D. (2017). Crossrail LFB Emergency Services Liaison. *Crossrail Ltd*. Available at:

https://learninglegacy.crossrail.co.uk/documents/crossrail-lfb-emergency-services-liaison/.

Bundgaard, K., Klazinga, D., and Visser, M. (2011). Traditional procurement methods are broken: Can early contractor involvement be the cure?. *Terra et Aqua*, *124*, 25–30.

Cai, L., and Zhu, Y. (2015). The challenges of data quality and data quality assessment in the Big Data era. *Data Science Journal*, *14*(2), 1–10. http://dx.doi.org/10.5334/dsj-2015-002

Caldas, C. H., and Soibelman, L. (2003). Automating hierarchical document classification for construction management information systems. *Automation in Construction*, *12*, 395– 406. https://doi.org/10.1016/S0926–5805(03)00004–9

Cao, D., Li, H., Wang, X. and Lu, W., (2017). Building Information Modelling (BIM) for Construction Safety Management. *Automation in Construction*, 73, pp.94-105.

Cao, M., Duan, X., and Zhao, X. (2019). Construction of Integrated Evaluation System for Green Port Construction Based on BIM Technology. *Journal of Coastal Research*, 1073–1079.

Caputo, A., and Kargina, M. (2021). A user-friendly method to merge Scopus and Web of Science data during bibliometric analysis. *Journal of Marketing Analytics*, 1–7.

Carvalho, J. P., Bragança, L., and Mateus, R. (2019). Optimising building sustainability assessment using BIM. Autom. Constr., 102, 170–182.

Catherine, W., and Alistair, W. (2018). *Crossrail's Information System*. Crossrail Ltd. Available at: https://learninglegacy.crossrail.co.uk/documents/crossrails-information-system/.

Ceccon, L., and Villa, D. (2021). AI-BIM Interdisciplinary Spill-Overs: Prospected Interplay of AI and BIM Development Paradigms. In C. Bolognesi and D. Villa (Eds.), From Building Information Modelling to Mixed Reality, *Springer Tracts in Civil Engineering* (pp. 195– 217). Cham: Springer International Publishing. https://doi.org/10.1007/978–3-030–49278–6_12

Chan, P. W., and Ejohwomu, O. (2018). How Does Project Management Relate to Productivity?: A Systematic Review of Published Evidence. *Association for Project Management*.

Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., and Wirth, R. (2000). *CRISP-DM 1.0: Step-by-step data mining guide*. SPSS Inc.

Chavarro, D., Ràfols, I., and Tang, P. (2018). To What Extent Is Inclusion in the Web of Science an Indicator of Journal 'Quality'? *Research Evaluation*, *27*, 106–118.

Chellappandi, P., and Vijayakumar, C. S. (2018). Bibliometrics, Scientometrics, Webometrics/Cybermetrics, Informetrics and Altmetrics--An Emerging Field in Library and Information Science Research. *Shanlax International Journal of Education*, *7*(1), 5–8

Chellappandi, P., and Vijayakumar, C. S. (2018). Bibliometrics, Scientometrics, and Informetrics: An Overview. In R. K. Sharma and A. K. Srivastava (Eds.), Scientometrics and Informetrics (pp. 1–13). Cham: Springer.

Chelson, D. E. (2010). The effects of building information modelling on construction site productivity (*Doctoral dissertation*).

Chen, M., Han, J., and Yu, P. S. (1996). Data mining: An overview from a database perspective. *IEEE Transactions on Knowledge and Data Engineering*, 8(6), 866-883.

Chen, Q. (2020). The Influence of Marine Industry Agglomeration on Regional Economic.

Chen, Q., de Soto, B. G., and Adey, B. T. (2018). Construction automation: Research areas, industry concerns and suggestions for advancement. *Automation in Construction*, *94*, 22–38.

Chen, Q., Sun, S., and Deng, M. (2015). Design and Construction of the Provincial Geo-Spatial Information Sharing Platform in China. *Journal of Coastal Research*, *73*.

Chen, W., Wei, Z., Zhang, H., Wang, J., and Miao, Z. (2019). Application of BIM technology in assembly bridge project. *Beijing Jiaotong Daxue Xuebao/Journal of Beijing Jiaotong University*, *43*, 65–70. https://doi.org/10.11860/j.issn.1673–0291.20190038

Chen, Y., Nicholson, D. P., Ingram, P., Hardy, S., Liew, H. L., Farooq, I., and Biscontin, G. (2015). Application of the Observational Method in Crossrail Projects. *Crossrail Ltd*. Available at: https://learninglegacy.crossrail.co.uk/documents/application-observational-method-crossrail-projects/.

Chen, Z., and Bareka, A. (2022)~~.~~ A text mining approach to performance enhancement in BIM pervasive major project delivery. *In International Post Graduate Research Conference (IPGRC)* 2022: Resilience in Research and Practice.

Chen, Z., (2019). Grand challenges in construction management. *Frontiers in Built Environment*, *5*, p.31.

Chen, Z., Agapiou, A., and Li, H. (2020). A Benefits Prioritization Analysis on Adopting BIM Systems Against Major Challenges in Megaproject Delivery. *Frontiers in Built Environment*, *6*, 26. https://doi.org/10.3389/fbuil.2020.00026

Chen Z., Agapiou A., Li, H., and Xu Q (2021) A TRIZ Approach to Reliable Megaproject Sustainability. *Frontiers in Built Environment,* 7:650699. DOI: 10.3389/fbuil.2021.650699.

Cheng, J. C. P., Chen, W., Chen, K., and Wang, Q. (2020). Data-driven predictive maintenance planning framework for MEP components based on BIM and IoT using machine learning algorithms. *Automation in Construction*, *112*, Article 103087. https://doi.org/10.1016/j.autcon.2020.103087

Cheng, L. (2017). Optimization Model of BIM-based Three-dimensional Reconstruction Technology and Engineering Model of Visual Perception. *In* *international symposium 2017 - mechanical and electronical systems and control engineering*.

Cheng, M., Mikulski, P. (2014). Ground settlement behaviour in chalk due to TBM excavations. ICE Publishing. Available at: https://learninglegacy.crossrail.co.uk/documents/ground-settlement-behaviour-in-chalk-due-to-tbm-excavations/.

Cheng, M.-Y., Tsai, H.-C., and Hsieh, W.-S. (2009). Web-based conceptual cost estimates for construction projects using Evolutionary Fuzzy Neural Inference Model. *Automation in Construction*, *18*, 164–172. https://doi.org/10.1016/j.autcon.2008.07.001

Chinowsky, P., Diekmann, J., and Galotti, V. (2008). Social Network Model of Construction. *Journal of Construction Engineering and Management*, *134*(10), 804–812. https://doi.org/10.1061/(ASCE)0733–9364(2008)134:10(804)

Choo, S., Lee, J., and Kim, K. (2019). Semantic interoperability framework for BIM-based integrated building data. *Journal of Computing in Civil Engineering*, 33(2), pp.04018061.

Choo, S., Park, H., Kim, T., and Seo, J. (2019). Analysis of trends in Korean BIM research. *Automation in Construction*, *99*, 111–119. https://doi.org/10.1016/j.autcon.2018.11.017

CIOB Planning protocol 2021 (CIOB PP21) (2023) *CIOB Academy*. Available at: https://www.ciobacademy.org/product/ciob-planning-protocol-2021-ciob-pp21/ (Accessed: 11 August 2023).

Cleves, J., and Dal Gallo, L. (2012). Integrated Project Delivery: The Game Changer. In *Proceedings of the American Bar's Construction Law Forum 2012*, Las Vegas (USA).

Coates, P., Arayici, Y., Koskela, K., Kagioglou, M., Usher, C., and O'Reilly, K. (2010). The key performance indicators of the BIM implementation process. *Journal of Facilities Management*, *8*(4), 267–282. https://doi.org/10.1108/14725961011087156

Cobbing, C., Webb, J., and Bewes, O. (2018). Control of Railway Induced Groundborne Noise and Vibration from the UK's Crossrail Project. *Crossrail Ltd*, ICE Publishing. Available at: https://learninglegacy.crossrail.co.uk/documents/control-of-railway-induced-groundborne-noise-and-vibration-from-the-uks-crossrail-project/.

Cohen, A. M., Ambert, K., and McDonagh, M. (2009). Cross-Topic Learning for Work Prioritization in Systematic Review Creation and Update. *Journal of the American Medical Informatics Association*, *16*(5), 690–704. https://doi.org/10.1197/jamia.M3162

Connolly-Barker, M., Kliestik, T., Suler, P., and Zvarikova, K. (2020). Real-time decision- making in the information technology-driven economy: Planning, managing, and operating smart sustainable cities. *Geopolitics, History and International Relations*, 12(1), 73–79.

Consterdine, R., and King, J. (2018). Maintenance and Cost Analysis on Major Subsurface Rail Projects – Crossrail. *Crossrail Ltd*, ICE Publishing. Available at: https://learninglegacy.crossrail.co.uk/documents/maintenance-and-cost-analysis-on-major-subsurface-rail-projects-crossrail/.

Consterdine, R., Halliday, B., King, J., Birch, T., and Dubery, P. (2021). Managing Reliability Growth in Practice. *Crossrail Ltd*, ICE Publishing. Available at: https://learninglegacy.crossrail.co.uk/documents/managing-reliability-growth-in-practice/.

Cox, R. F., Issa, R. R. A., and Ahrens, D. (2003). Management's Perception of Key Performance Indicators for Construction. *Journal of Construction Engineering and Management*, *129*(2), 142–151. https://doi.org/10.1061/(ASCE)0733–9364(2003)129:2(142)

Cresswell, J.W. and Plano Clark, V.L. (2007) *Designing and Conducting Mixed Methods Research*. Thousand Oaks, CA: Sage Publications.

Creswell, J.W. (2014) Research Design: *Qualitative, Quantitative, and Mixed Methods Approaches*. 4th ed. Thousand Oaks, CA: SAGE Publications.

Creswell, J.W., and Clark, V.L.P. (2011) *Designing and Conducting Mixed Methods Research*. 2nd ed. Thousand Oaks, CA: SAGE Publications.

Criminale, A., and Langar, S. (2017). Challenges with BIM Implementation: A Review of the Academic Literature. *Journal of Information Technology in Construction*, *22*, 412–429. https://doi.org/10.36680/j.itcon.2017.022

Cross, R. L., and Parker, A. (2004). The hidden power of social networks: Understanding how work really gets done in organizations. *Harvard Business School Press*, Boston, Mass.

Crossrail Ltd. (2017). Crossrail BIM principles. *ICE Publishing*. Available at: https://learninglegacy.crossrail.co.uk/documents/crossrail-bim-principles/.

Crossrail Ltd. (2018). Crossrail asset information guide. *Crossrail Ltd*. Available at: https://learninglegacy.crossrail.co.uk/documents/crossrail-asset-information-guide/.

Crossrail Ltd. (2022). Crossrail learning legacy. Available online at: https://learninglegacy.crossrail.co.uk/.

Crotty, M. (1998) *The Foundations of Social Research: Meaning and Perspective in the Research Process*. London: SAGE Publications.

Crotty, M. (2003) *The Foundations of Social Research*: Meaning and Perspective in the Research Process. London: SAGE Publications.

Csardi, G. and Nepusz, T., (2006). The igraph software package for complex network research. *InterJournal Complex Systems*, 1695, pp.1-9.

Dahlqvist, F., Green, A., Morgan, P., and Palter, R. (2020). High Stakes: How Investors Can Manage Risk in The New Infrastructure Environment. [online] *Global Infrastructure Initiative*. https://www.globalinfrastructureinitiative.com/article/high- stakes-how-investors-can-manage-risk-new-infrastructure-environment.

Dai, J., Goodrum, P. M., Maloney, W. F. (2009). Construction Craft Workers' Perceptions of the Factors Affecting Their Productivity. *J. Constr. Eng. Manage*. 135, 217–226. https://doi.org/10.1061/(ASCE)0733–9364(2009)135:3(217)

Dammann, O. (2019). Data, Information, Evidence, and Knowledge: A Proposal for Health Development. *Journal of Coastal Research*, 263–267.

Dave, B., Buda, A., Nurminen, A., Främling, K. (2018). A framework for integrating BIM and IoT through open standards. *Automation in Construction*, *95*, 35–45. https://doi.org/10.1016/j.autcon.2018.07.022

Davenport, T.H. and Ronanki, R. (2018) 'Artificial Intelligence for the Real World.' *Harvard Business Review*, 96(1), pp. 108–116.

David, M. (2017). Will rooms operate on loss of mains power? *Crossrail Ltd*. Available at: https://learninglegacy.crossrail.co.uk/documents/will-rooms-operate-loss-mains-power/.

De Boissieu, A. (2022). Introduction to computational design: Subsets, challenges in practice and emerging roles. In M. Bolpagni, R. Gavina, and D. Ribeiro (Eds.), *Industry 4.0 for the built environment*. New York, NY: Springer. https://hdl.handle.net/2268/262318

Dell'Acqua, G., De Oliveira, S. G., Biancardo, S. A. (2018). Railway-BIM: Analytical review, data standard and overall perspective. *Ingegneria Ferroviaria*, *73*, 901–923.

Dentten, R. (2018). Better Information Management to Optimise Whole Life Business Decisions. *Crossrail Ltd*, ICE Publishing. Available at: https://learninglegacy.crossrail.co.uk/documents/better-information-management-to-optimise-whole-life-business-decisions/.

Department for Levelling Up, Housing and Communities. (2021) Approved documents. GOV.UK. https://www.gov.uk/government/collections/approved-documents (Accessed: January 22, 2023).

DfLU. (2021). Department for Levelling Up, Housing and Communities. UK Government. Available at: https://www.gov.uk/government/organisations/department-for-levelling-up-housing-and-communities.

Dorigo, M., and Stützle, T. (2010). *Ant colony optimization: Overview and recent advances* (pp. 311–351). Cham: Springer International Publishing.

Dorji, S., and Kirikova, M. (2012). Data, information, and knowledge modelling in worksystem networks. *BIR Workshops*. https://ceur-ws.org/Vol-1684/paper22.pdf

Dowden, O. (2022). *National Data Strategy*. [online] GOV.UK. Available at: https://www.gov.uk/guidance/national-data-strategy

Dozzi, S. P., and AbouRizk, S. M. (1993). Productivity in construction. Ottawa: Institute for Research in Construction, *National Research Council*.

Durdyev, S., and Mbachu, J. (2011) On-site labour productivity of New Zealand construction industry: Key constraints and improvement measures. *Construction Economics and Building*, *11*(3), 18–33.

Eastman, C. M., Eastman, C., Teicholz, P., Sacks, R., and Liston, K. (2011). BIM handbook: A guide to building information modelling for owners, managers, designers, engineers and contractors. London: John Wiley and Sons.

Ebert, C., Duarte, C. H. C. (2018). Digital Transformation. *IEEE Softw*. 35, 16–21. https://doi.org/10.1109/MS.2018.2801537

Echchakoui, S. (2020). Why and how to merge Scopus and Web of Science during bibliometric analysis: The case of sales force literature from 1912 to 2019. *Journal of Marketing Analytics*, 1–20.

Eckblad, S., Ashcraft, H., Audsley, P., Blieman, D., Bedrick, J., Brewis, C., Hartung, R. J., Onuma, K., Rubel, Z., and Stephens, N. D. (2007). Integrated Project Delivery - A Working Definition. *AIA California Counci*l, Sacramento, CA.

Education. (2018). Efficiency in BIM-Based Collaborative Design: An Empirical Study in China. *Buildings*, *11*, 486.

Ehrbar, H. (2017). BIM in infrastructure construction of German Railway - prerequisites for a successful introduction. *Bautechnik*, https://doi.org/10.1002/bate.201700012

Eisenhardt, K. M. (1989). Building Theories from Case Study Research. Academy of Management Review, 14(4), 532–550.

Eisenhardt, K.M. (1989) *Building theories from case study research*, *Academy of Management Review*, 14(4), pp. 532–550. https://doi.org/10.5465/amr.1989.4308385.

Eisenmann, S., and Park, B. (2012). Building information modelling's impact on team performance. *The Associated Schools of Construction*.

El-Diraby, T., Krijnen, T. and Papagelis, M., (2017). BIM-based collaborative design and socio-technical analytics of green buildings. *Automation in Construction*, *82*, pp.59-74.

El-Gohary, K. M., and Aziz, R. F. (2014). Factors influencing construction labor productivity in Egypt. *Journal of Management in Engineering*, *30*, 1–9. https://doi.org/10.1061/(ASCE)ME.1943–5479.0000168

Elliot, J. (2016). Works Information Study (Supply Chain Quality Requirements). *Crossrail Ltd*. Available at: https://learninglegacy.crossrail.co.uk/document/supply-chain-quality-requirements/.

Enshassi, A., Mohamed, S., and Mayer, P. E. (2007). Factors affecting labour productivity in building projects in the Gaza Strip. *Journal of Civil Engineering and Management*, *13*, 245–254. https://doi.org/10.3846/13923730.2007.9636444

Erdogan, B., Abbott, C., and Aouad, G. (2010). Construction in year 2030: Developing an integrated future. *Journal of Information Technology in Construction*, *15*, 411–425.

Erikawati, D. (2023) 'Realism and Positivism Paradigm in Research', *Journal of Social Sciences Research,* 10(2), pp. 13-25.

Evans, A. (2014). The Protection of the 400kV Cables at Pudding Mill Lane. *Crossrail Ltd*. Available at: https://learninglegacy.crossrail.co.uk/documents/the-protection-of-the-400kv-cables-at-pudding-mill-lane/.

Eweje, J., Turner, R., and Müller, R. (2012). Maximizing strategic value from megaprojects: The influence of information-feed on decision-making by the project manager. *International Journal of Project Management*, *30,* 639–651. https://doi.org/10.1016/j.ijproman.2012.01.004

Farooq, I., Place, D., Steele, B., Liew, H. L., and Vitorino, S. (2016). An innovative verification process speeds construction of Crossrail's Moorgate shaft. *Crossrail Ltd*. Available at: https://learninglegacy.crossrail.co.uk/documents/innovative-verification-process-speeds-construction-crossrails-moorgate-shaft/.

Feldman, R., and Sanger, J. (2006). The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data. *Cambridge University Press*. https://doi.org/10.1017/CBO9780511546914

Fellows, R. and Liu, A. (2008) *Research Methods for Construction*. 3rd ed. Oxford: Wiley-Blackwell.

Fennell, A. (2020). Digital Groundwork: How UK Construction Firms Can Close the Productivity Gap. [Blog] *Autodesk Construction Cloud*, https://constructionblog.autodesk.com/uk-construction/

Fensel, D., Şimşek, U., Angele, K., Huaman, E., Kärle, E., Panasiuk, O., Toma, I., Umbrich, J., Wahler, A., (2020). Introduction: What Is a Knowledge Graph?, in: Knowledge Graphs: Methodology, Tools and Selected Use Cases. *Springer International Publishing, Cham*, pp. 1–10. https://doi.org/10.1007/978-3-030-37439-6_1

Fischer, M. (2013). 2nd Annual BIM conference in the Middle East. In Leasons Learnt from Hundreds of BIM implementations. *Abu Dhabi*.

Flyvbjerg, B. (2006) *Five Misunderstandings About Case-Study Research, Qualitative Inquiry,* 12(2), pp. 219-245. https://doi.org/10.1177/1077800405284363.

Foxe, D. M. (2010). Building Information Modelling for Constructing the Past and Its Future. APT Bulletin: *The Journal of Preservation Technology*, *41*(4), 39–45.

Frances, O., and Hayman-Joyce, R. (2015). Compensation grouting. *Crossrail Ltd*. Available at: https://learninglegacy.crossrail.co.uk/documents/compensation-grouting/.

Freeman, R. (2008). Labour productivity indicators. Organization for Economic and Cooperation and Development (OECD).

Frey, L., Botan, C., and Kreps, G. (1999). Investigating communication: An introduction to research methods. (2nd ed.) Boston: Allyn and Bacon.

Fung, W. (2017). Application of stair pressurisation in Crossrail stations, shafts, and portals. *Crossrail Ltd*. Available at: https://learninglegacy.crossrail.co.uk/documents/application-stair-pressurisation-crossrail-stations-shafts-portals/.

Fung, W. (2018). Implementation of Energy Saving Initiatives in Design Review. *Crossrail Ltd*, ICE Publishing. Available at: https://learninglegacy.crossrail.co.uk/documents/implementation-energy-saving-initiatives-design-review/.

Fung, W., Shone, H., and Ash, A. (2017). Use of non-metallic materials in ventilation airways: A review of aerodynamic performance and material requirements. *Crossrail Ltd*. Available at: https://learninglegacy.crossrail.co.uk/documents/use-of-non-metallic-materials-in-ventilation-airways-a-review-of-aerodynamic-performance-and-material-requirements/.

Gadd, K. (2011) *TRIZ for engineers: Enabling inventive problem solving*. London: John Wiley and Sons.

Galkin, M., Auer, S., Vidal, M.E. and Scerri, S. (2017), April. Enterprise knowledge graphs: A semantic approach for knowledge management in the next generation of enterprise information systems. In *International Conference on Enterprise Information Systems* (Vol. 2, pp. 88-98). SCITEPRESS.

Ganbat, T., Chong, H. Y., Liao, P. C., and Wu, Y. D. (2018). A bibliometric review on risk management and building information modelling for international construction. *Advances in Civil Engineering*, *2018*.

Gercek, B., Tokdemir, O., Ilal, M., and Gunaydin, H. (2017). BIM execution process of construction companies for building projects. *Resileint structures and sustainable construction*. ISEC Press.

Gerring, J. (2004) What is a case study and what is it good for*? American Political Science Review*, 98(2), pp. 341-354. https://doi.org/10.1017/S0003055404001182.

Ghaffarianhoseini, A., Tookey, J., Ghaffarianhoseini, A., Naismith, N., Azhar, S., Efimova, O., and Raahemifar, K. (2017). Building Information Modelling (BIM) uptake: Clear benefits, understanding its implementation, risks and challenges. *Renewable and Sustainable Energy Reviews*, *75*, 1046–1053.

Ghosh, A.; Edwards, D. J.; and Hosseini, M.R. (2020). Patterns and trends in Internet of Things (IoT) research: Future applications in the construction industry. *Engineering, construction and architectural management*, *28*(2), 457–481.

Giani, R., Piva, D., and Benassi, R. (2020). Integrated BIM-GIS and IoT systems for tunnel maintenance. *Tunnelling and Underground Space Technology*, 96, p.103234.

Gibbons, M., Limoges, C., Nowotny, H., Schwartzman, S., Scott, P., and Trow, M. (1994) *The New Production of Knowledge: The Dynamics of Science and Research in Contemporary Societies*. London: SAGE Publications.

Giorgi, F. M., Ceraolo, C., and Mercatelli, D. (2022). The R language: An engine for bioinformatics and data science. Life, *12*(5), 648.

Giorgi, S., Flesca, S., and Lodi, P. (2022). A review of text mining techniques for big data analytics. *Information Systems Frontiers*, 24(3), 755-775.

Glegg, S. M. N., Jenkins, E., and Kothari, A., (2019). How the study of networks informs knowledge translation and implementation: A scoping review. *Implementation Sci*, *14*, 34. https://doi.org/10.1186/s13012-019-0879-1

Goldberg, D. E. (1989). Genetic algorithms in search, optimization, and machine learning (Vol. 412). Addison-Wesley.

Goldkuhl, G. (2012) ‘Pragmatism vs. Interpretivism in Qualitative Information Systems Research’, *European Journal of Information Systems*, 21(2), pp. 135-146.

Goodrum, P. M., Haas, C. T. (2002). Partial Factor Productivity and Equipment Technology Change at Activity Level in U.S. Construction Industry. *J. Constr. Eng*. Manage. 128, 463–472. https://doi.org/10.1061/(ASCE)0733–9364(2002)128:6(463)

Gopalakrishnan, B., Arunachalam, K., and Girija, S. (2017). Big Data issues in construction: A systematic review. *Automation in Construction*, 81, pp.105-116.

Gopalakrishnan, K., Agrawal, A., and Choudhary, A. (2017). Big Data in building information modelling research: Survey and exploratory text mining. *MOJ Civil Eng*, *3*(6), Article 00087.

Government Digital Services (2017). *Digital Service Standard*. [online] Available at: https://www.gov.uk/service-manual/service-standard.

Gray, D.E. (2013) *Doing Research in the Real World*. 3rd edn. London: Sage Publications.

Grobelnik, M., Mladenic, D., and Jermol, M. (2002). Exploiting text mining in publishing and education. In *Proceedings of the ICML-2002 workshop on data mining lessons learned* (pp. 34–39).

Guba, E.G. and Lincoln, Y.S. (2005) *Paradigmatic controversies, contradictions, and emerging confluences*. In: Denzin, N.K. and Lincoln, Y.S. (eds.) *The SAGE Handbook of Qualitative Research*. 3rd ed. Thousand Oaks, CA: SAGE Publications, pp.191-215.

Guo, F., Zhou, J., Zhang, W., Ding, L. and Luo, H., (2022). A comprehensive review of Building Information Modeling (BIM) in the architecture, engineering and construction (AEC) industry: Advances, applications, and future trends. *Automation in Construction*, 135, p.104098.

Guo, S., Ding, Y., Li, H., and Lu, W. (2022). BIM-based application in the AEC industry: The latest trends, gaps, and future directions. *Automation in Construction*, 133, p.104038.

Guo, X., Li, Z., and Xie, X. (2022). Big Data in construction: A survey. *Automation in Construction*, 128, 103737.

Guo, Y., Huang, Z.-L., Guo, J., Li, H., Guo, X.-R., and Nkeli, M. J. (2019). Bibliometric Analysis on Smart Cities Research. *Sustainability*.

Habibi, M., Kermanshachi, S. and Rouhanizadeh, B., (2019). Identifying and measuring engineering, procurement, and construction (EPC) key performance indicators and management strategies. *Infrastructures*, *4*(2), p.14.

Hadzaman, N. A. H., Takim, R., Mohammad, A.-H. N. M. F. (2016). An exploratory study: Building information modelling execution plan (BEP) procedure in mega construction projects. *Malaysian Construction Research Journal*, *18*, 29–40.

Haeussler, M., Borrmann, A. (2020). Model-based quality assurance in railway infrastructure planning. *Automation in Construction*. https://doi.org/10.1016/j.autcon.2019.102971

Hall, J.W. (2012) *Research paradigms and meaning making: A primer*. *International Journal of Qualitative Methods*, 11(4), pp. 375-394.

Halligan, D. W., Demsetz, L. A., Brown, J. D., Pace, C. B. (1994). Action-Response Model and Loss of Productivity in Construction. *Journal of Construction Engineering and Management*, *120*, 47–64. https://doi.org/10.1061/(ASCE)0733–9364(1994)120:1(47)

Hamma-adama, M., and Kouider, T. (2019). Comparative Analysis of BIM Adoption Efforts by Developed Countries as Precedent for New Adopter Countries. *Current Journal of Applied Science and Technology*.

Han, Y. (2019). Issues on Intelligent Railways Technology by Combining Spatial Ontology and Engineering Cybernetics. *Journal of Railway Engineering Society*, *36*, 49–53.

Hanisch, B., and Wald, A. E. (2012). A Bibliometric View on the Use of Contingency Theory in Project Management Research. *Project Management Journal*, *43*, 23–24.

Hao, T., Chen, X., Li, G., and Yan, J. (2018). A bibliometric analysis of text mining in medical research. *Soft Computing*, *22*(23), 7875–7892. https://doi.org/10.1007/s00500-018-3511-4

Haq, M. (2014) ‘Exploring Ontology, Epistemology, and Methodology in Management Research’, *Journal of Research in Marketing and Entrepreneurship*, 6(1), pp. 56-69.

Harding, J. A., Shahbaz, M., Srinivas, Kusiak, A. (2006). Data Mining in Manufacturing: A Review. *Journal of Manufacturing Science and Engineering*, *128*, 969–976. https://doi.org/10.1115/1.2194554

Harpaz, R., Callahan, A., Tamang, S., Low, Y., Odgers, D., Finlayson, S., Jung, K., LePendu, P., and Shah, N. H. (2014). Text mining for adverse drug events: The promise, challenges, and state of the art. Drug safety, *37*, 777–790.

Harpaz, R., Tapiero, C., and Li, Y. (2014). Text mining and the challenges of unstructured data. *International Journal of Information Management*, 34(2), 116-124. https://doi.org/10.1016/j.ijinfomgt.2013.12.004

Harris, C. (2014). The impact of text mining on business decision-making: A growing trend. *International Journal of Business Intelligence and Data Mining*, 10(1), 1-11. https://doi.org/10.1504/IJBIDM.2014.061633

Hassan Ibrahim, N. (2013). Reviewing the evidence: USE of digital collaboration technologies in major building and infrastructure projects. *Journal of Information Technology in Construction* 18, 40–63.

Hassani, H., Beneki, C., Unger, S., Mazinani, M. T., and Yeganegi, M. R. (2020). Text mining in Big Data analytics. Big Data and Cognitive Computing, *4*(1), p.1.

Hassani, H., Silva, E. S., and Ghodsi, M. (2020). The challenges of big data processing in text mining. *Journal of Big Data*, 7(1), 15-32. https://doi.org/10.1186/s40537-019-0206-3

Hegazy, T., Fazio, P., Moselhi, O. (1994). Developing Practical Neural Network Applications Using Back-Propagation. *Computer-Aided Civil and Infrastructure Engineering*, *9*, 145–159. https://doi.org/10.1111/j.1467–8667.1994.tb00369.x

Heinzelmann, C., Bödefeld, J., Westendarp, A. (2018). Challenges in waterways engineering - From the Federal Transport Infrastructure Plan to Building Information Modelling. *Bautechnik*, *95*, 499–504. https://doi.org/10.1002/bate.201800027

Herbsman, Z., Ellis, R. (1990). Research of factors influencing construction productivity. *Construction Management and Economics*, *8*, 49–61. https://doi.org/10.1080/01446199000000005

Hillier, L. (2016). *Young Crossrail Programme*. Case Study. Published 27 September 2016. Available at: https://learninglegacy.crossrail.co.uk/documents/young-crossrail-programme/

HM Government. (2011.) *Government construction strategy 2011-2015*. [online] Available at: https://www.gov.uk/government/publications/government-construction-strategy-2011-2015

HM Government. (2016). *Government construction strategy 2016-2020*. [online] Available at: https://www.gov.uk/government/publications/government-construction-strategy-2016-2020

Hochmuth, M., Breinig, W. (2016). BIM pilot project Auenbach viaduct - Innovative design methods in bridge construction. *BAUTECHNIK*. https://doi.org/10.1002/bate.201600041

Hochmuth, M., Nguyen, T., Haeussler, M. (2020). Innovative and modern planning methods in bridge construction. *BAUTECHNIK*. https://doi.org/10.1002/bate.202000001

Hodorog, A., Petri, I., Rezgui, Y., and Hippolyte, J.-L. (2021). Building information modelling knowledge harvesting for energy efficiency in the construction industry. *Clean Technologies and Environmental Policy*, *23*, 1215–1231. https://doi.org/10.1007/s10098-020-02000-z

Holzer, D. (2015). BIM for procurement - Procuring for BIM. In: R H Crawford and A Stephan (Eds.) Proceedings of Living and Learning: Research for a Better Built Environment: *49th International Conference of the Architectural Science Association 2015*, Melbourne. The Architectural Science Association and The University of Melbourne, 237–246.

Holzer, D. (2015). *The BIM manager's handbook: Guidance for professionals in architecture, engineering, and construction*. London: John Wiley and Sons.

Honnibal, M., and Montani, I. (2017). spaCy 2: Natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing.

Hoppe, S., Schumann, R. (2018). BIM tool for dam surveillance. *WasserWirtschaft*, *108,* 48–51. https://doi.org/10.1007/s35147–018–0206–9

Hosch, W. (2022). genetic algorithm | *computer science*. [online] Encyclopedia Britannica. https://www.britannica.com/technology/genetic-algorithm

Hosseini, M. R., Becerik-Gerber, B., and Lavy, S. (2017). BIM for decision-making in construction project management: A review and a case study. *Automation in Construction*, 79, 93-106.

Hosseini, M. R., Maghrebi, M., Akbarnezhad, A., Martek, I., Arashpour, M. (2018a). Analysis of Citation Networks in Building Information Modelling Research. *J. Constr. Eng*. Manage. 144, 04018064. https://doi.org/10.1061/(ASCE)CO.1943–7862.0001492

Hosseini, M. R., Martek, I., Papadonikolaki, E., Sheikhkhoshkar, M., Banihashemi, S., Arashpour, M. (2018b). Viability of the BIM Manager Enduring as a Distinct Role: Association Rule Mining of Job Advertisements. *J. Constr. Eng. Manage*. 144, 04018085. https://doi.org/10.1061/(ASCE)CO.1943–7862.0001542

Hotho, A., Nürnberger, A., and Paass, G. (2005). A Brief Survey of Text Mining. *LDV Forum*, 20, 19–62.

Hou, P., and Wu, J. (2019). Construction of Path Selection Model for Economic Development of Coastal Rural Leisure Industry. *Journal of Coastal Research*, *843*–849.

Howard, R., Björk, B.-C. (2008). Building information modelling – Experts' views on standardisation and industry deployment. *Advanced Engineering Informatics*, *22*, 271–280. https://doi.org/10.1016/j.aei.2007.03.001

Huang, C., Cai, Y., and Chen, C. (2020). Situation and Countermeasure of Free Ports and Marine Strategic Emerging Industries: A Case of Zhoushan in China. *Journal of Coastal Research*, 252–257.

Huang, J., Mao, Y., Wang, D., Zhu, J., Zhou, Y. (2019). Construction schedule-cost joint control of RCC dams based on BIM. *Advances in Science and Technology of Water Resources*, *39*, 66–72 and 88. https://doi.org/10.3880/j.issn.1006–7647.2019.05.010

Huang, Y., Wang, S., Zhao, J. (2017). The Application Research on BIM Technology in Road and Bridge Projects in the Western Yunnan Area. *AGRO FOOD INDUSTRY HI-TECH*.

Hung, J. L. (2012). Trends of e-learning research from 2000 to 2008: Use of text mining and bibliometrics. *British journal of educational technology, 43*(1), 5–16.

Ibbs, W., Nguyen, L. D., Lee, S. (2007). Quantified Impacts of Project Change. *J. Prof. Issues Eng. Educ*. Pract. *133*, 45–52. https://doi.org/10.1061/(ASCE)1052–3928(2007)133:1(45).

IBM. (2024). *Artificial intelligence*. [online] Available at: https://www.ibm.com/topics/artificial-intelligence

Ibrahim, N. (2013). Reviewing the evidence: USE of digital collaboration technologies in major building and infrastructure projects. *Journal of Information Technology in Construction*, *18*, 40–63.

IGI Global. (2022). *What is fuzzy system?*. https://www.igi-global.com/dictionary/fuzzy-system/11786

Ignatow, G., and Mihalcea, R. (2016). Text mining: A guidebook for the social sciences. Sage Publications.

Industry, G. (2016). 4.0 Survey: What we mean by Industry 4.0/Survey key findings. Blueprint for digital success, Industry, 4.

Irani, Z., Kamal, M. M. (2014). Intelligent Systems Research in the Construction Industry. *Expert Systems with Applications*, *41*, 934–950. https://doi.org/10.1016/j.eswa.2013.06.061

Irwin, D. (2016). The Geo-centric Railway - Why Location Matters in the Rail Industry. *Crossrail Ltd*. Available at: https://learninglegacy.crossrail.co.uk/wp-content/uploads/2018/05/12D-004_The-Geo-centric-Railway.pdf.

Irwin, D., and Tamash, N. (2016). Building a spatial infrastructure for Crossrail. *Crossrail Ltd*. Available at: https://learninglegacy.crossrail.co.uk/document/building-a-spatial-infrastructure-for-crossrail/.

ISO (2021) ISO 19475:2021 Document management — Minimum requirements for the storage of documents. International Organization for Standardization (ISO), Geneva, Switzerland.

Ittoo, A., Ferjani, A., Ben Arfa, N. and Chraibi, M. (2016). Data Integration in BIM Projects: Challenges and Solutions. *Automation in Construction*, 72, pp. 13-28.

Ittoo, A., Law, P., and Pinto, H. (2016). *Real-time analytics for big data processing: Challenges and solutions*. Springer.

Ittoo, A., Parvez, Z., and Abid, A. (2016). Text mining applications in various industries. *International Journal of Computer Applications*, 139(2), 23-28. https://doi.org/10.5120/ijca2016908392

Jackson, R. H. G., and Wood, A. (2013). The performance of insolvency prediction and credit risk models in the UK: A comparative study. *The British Accounting Review*, *45*, 183–202. https://doi.org/10.1016/j.bar.2013.06.009

Jamal, K. A. A., Mohammad, M. F., Hashim, N., Mohamed, M. R., and Ramli, M. A. (2019). Challenges of Building Information Modelling (BIM) from the Malaysian architect's perspective. *In MATEC web of conferences* (Vol. 266, p. 05003). EDP Sciences.

Janssen, P., Chen, K. W., and Mohanty, A. (2016). Automated generation of BIM models. *In Proceedings of the 34th Education and research in Computer Aided Architectural Design in Europe* (eCAADe) Conference (pp. 583–590).

Järvisalo, J., Saris, N. E. (1975). Action of propranolol on mitochondrial functions--effects on energized ion fluxes in the presence of valinomycin. *Biochem. Pharmacol*. 24, 1701–1705. https://doi.org/10.1016/0006–2952(75)90009-x

Jergeas, G. F. (2009). Improving Construction Productivity on Alberta Oil and Gas Capital Projects. *Report to Alberta Finance and Enterprise Project Report*, Alberta.

Jiang, S., Zhang, H., and Zhang, J. (2013). Research on BIM-based Construction Domain Text Information Management. *J. Networks*, *8(*6), 1455–1464.

Jiang, S., Zhong, R., and Ding, L. (2016). Multilingual processing and development in AEC industry: Insights into BIM-based applications. *Journal of Computing in Civil Engineering*, 30(6), pp.04016025.

Jiedeerbieke, M., Niu, Z., Li, P., Zhang, Y., Kuai, P., Li, J. (2020). Dam safety visual monitoring system based on BIM technology and B/S architecture. *Journal of Drainage and Irrigation Machinery Engineering*, *38*, 583–588. https://doi.org/10.3969/j.issn.1674–8530.18.0226

Jifa, G. (2013). Data, information, knowledge, wisdom and meta-synthesis of wisdowm- coment on wisdom global and wisdowm cities. *Procedia computer science, 17*, 713- 719.

Jim, D., and Jorrin, S. (2020). *Benefits of Contractor Prototyping: GFRC Cladding on Crossrail*. Crossrail Ltd. Available at: https://learninglegacy.crossrail.co.uk/documents/benefits-of-contractor-prototyping-gfrc-cladding-on-crossrail.

Jin, R., Hancock, C. M., Tang, L., Wanatowski, D. (2017). BIM Investment, Returns, and Risks in China's AEC Industries. *J. Constr. Eng. Manage*. 143, 04017089. https://doi.org/10.1061/(ASCE)CO.1943–7862.0001408

Jing, P. (2017). The Application of BIM Technology in Simulation Modelling of Concrete Dam Construction. *AGRO Food Industry Hi-Tech*.

Jones, B. (2014). Integrated Project Delivery (IPD) for Maximizing Design and Construction Considerations Regarding Sustainability. *Procedia Engineering, 95*, 528–538. https://doi.org/10.1016/j.proeng.2014.12.214

Jones, Q., Ravid, G., and Rafaeli, S. (2004). Information overload and the message dynamics of online interaction spaces: A theoretical model and empirical exploration. *Information Systems Research*, *15*(2), 194–210.

Jones, S. A., and Laquidara-Carr, D. (2017). The Business Value of BIM for Infrastructure, *SmartMarket Report*. In Rep. to Dodge Data and Analytics, Balford, MA.

Juneja, A., and Das, N. N. (2019), February. Big Data quality framework: Pre-processing data in weather monitoring application. In 2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon) (pp. 559–563). IEEE.

Juneja, P., Kaur, H., and Gupta, A. (2019). Pre-processing techniques for text mining: A comprehensive survey. *International Journal of Computer Science and Information Securit*y, 17(9), 107-114. https://doi.org/10.2139/ssrn.3458432

Junior, Giuseppe and Pellanda, Paulo and Reis, Marcelo. (2019). Implementation Framework for BIM Adoption and Project Management in Public Organizations. https://doi.org/10.22260/ISARC2019/0016

Jurafsky, D., and Martin, J. H. (2019). *Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition* (3rd ed.). Pearson.

Kaewunruen, S., Lian, Q. (2019.) Digital twin aided sustainability-based lifecycle management for railway turnout systems. *Journal of Cleaner Production*. https://doi.org/10.1016/j.jclepro.2019.04.156

Kaewunruen, S., Sresakoolchai, J., Zhou, Z. (2020). Sustainability-Based Lifecycle Management for Bridge Infrastructure Using 6D BIM. *SUSTAINABILITY*. https://doi.org/10.3390/su12062436

Kakas, A., Cohn, D., Dasgupta, S., Barto, A., Carpenter, G., Grossberg, S., Webb, G., Dorigo, M., and Birattari, M. (2011). Ant colony optimization. Encyclopedia of machine learning. *Ant Colony Optimization: A Component-Wise Overview*; Martí, R., Ed, pp.1–28.

Kanrak, M., and Nguyen, H. O. (2022). Structure, characteristics and connectivity analysis of the Asian-Australasian cruise shipping network. *Maritime Policy and Management*, 49(6), 882–896.

Karaboga., D. (2005). An idea based on honey bee swarm for numerical optimization. *Technical Report TR06*, Erciyes University, Engineering Faculty, Computer Engineering Department.

Karan, P., and Irizarry, J. (2015). *BIM for project managers*. CRC Press.

Karaseva, M. (2024) *Research Approaches and Methodologies in IT Sector Analysis, Journal of Information Technology Studies*, 9(3), pp. 212-230.

Karlsson, I., and Ronndahl, C. (2018). A study of national BIM guidelines from around the world determining what future Swedish national BIM guidelines out to contain. https://www.divaportal.org/smash/get/diva2:1237159/FULLTEXT01.pdf

Kasravi, K. (2004). Improving the Engineering Processes with Text Mining, in: Volume 4: 24th Computers and Information in Engineering Conference. *Presented at the ASME 2004 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference,* *ASMEDC*, Salt Lake City, Utah, USA, pp. 1049–1051. https://doi.org/10.1115/DETC2004–57790

Kazaz, A., Manisali, E., Ulubeyli, S. (2008). Effect of basic motivational factors on construction workforce productivity in turkey. *Journal of civil engineering and management* 14, 95–106. https://doi.org/10.3846/1392–3730.2008.14.4

Kennedy, J., and Eberhart, R. (2010). Particle swarm optimization. In Encyclopedia of machine learning (pp. 760–766). Cham: Springer.

Kent, D. C., Becerik-Gerber, B. (2010). Understanding Construction Industry Experience and Attitudes toward Integrated Project Delivery. J. *Constr. Eng*. Manage. 136, 815–825. https://doi.org/10.1061/(ASCE)CO.1943–7862.0000188

Kenton, W. (2022). Least squares method definition. *Investopedia*. https://www.investopedia.com/terms/l/least-squares-method.asp

Khan, M. A., and Ahmed, A. (2012). Modelling Trade, Investment, Growth and Liberalisation: Case Study of Pakistan. *The Pakistan Development Review*, *51*(4), 187–208.

Khanzode, A., Fischer, M., and Reed, D. (2008). Benefits and lessons learned of implementing building virtual design and construction (VDC) technologies for coordination of mechanical, electrical, and plumbing (MEP) systems on a large healthcare project. *ITcon*, *13,* 324–342.

Kiburu, L., Njiraini, N., and Boso, N. (2023). Social networks and consumer technology usage: A systematic literature review and future research directions. *Cogent Business and Management*, *10*(1), p.2153487.

Kim, D.-G. (2018). Development of building information modelling (BIM) and mobile linking technology based on construction IT convergence technology. *Asia Life Sciences Supplement* 15, 2245–2258.

Kim, H., Shen, Z., Moon, H., Ju, K., Choi, W. (2016). Developing a 3D Intelligent Object Model for the Application of Construction Planning/Simulation in a Highway Project. *KSCE Journal of Civil Engineering 20*, 538–548. https://doi.org/10.1007/s12205–015–0463–4

Kim, Ki. (2015). Conceptual Building Information Modelling Framework for Whole-house Refurbishment based on LCC and LCA. *PhD thesis*, Aston University.

Kim, S., Chin, S., Han, J., Choi, C.-H. (2017). Measurement of Construction BIM Value Based on a Case Study of a Large-Scale Building Project. *J. Manage. Eng*. 33, 05017005. https://doi.org/10.1061/(ASCE)ME.1943–5479.0000551

Kim, S.H. and Kim, J. (2018). A text mining-based approach to categorizing knowledge for construction engineering projects. *Journal of Computing in Civil Engineering*, 32(1), pp.04017068.

King, J. (2018). Effective MEP Design Techniques and Strategy for Large Subsurface Metro Projects. *ICE Publishing*. Available at: https://learninglegacy.crossrail.co.uk/documents/effective-mep-design-techniques-strategy-large-subsurface-metro-projects/.

King, M. (2018). Protection Against Fire for the UK Crossrail Tunnel Structures. *Crossrail Ltd*, ICE Publishing. Available at: https://learninglegacy.crossrail.co.uk/documents/protection-against-fire-for-the-uk-crossrail-tunnel-structures/.

Knobe, J. and Nichols, S. (2017) *Experimental Philosophy*. Oxford: Oxford University Press.

Kolodner, J. L. (2014). *Case-based reasoning*. Morgan Kaufmann.

Konchar, M., Sanvido, V. (1998). Comparison of U.S. Project Delivery Systems. *Journal of Construction Engineering and Management*, *124*, 435–444. https://doi.org/10.1061/(ASCE)0733–9364(1998)124:6(435)

Koseoglu, O., Keskin, B., and Ozorhon, B. (2019). Challenges and enablers in BIM-enabled digital transformation in mega projects: The Istanbul new airport project case study. *Buildings*, *9*(5), 115.

Koseoglu, O., Sakin, M., Arayici, Y. (2018). Exploring the BIM and lean synergies in the Istanbul Grand Airport construction project. *Engineering Construction and Architectural Management*. https://doi.org/10.1108/ECAM-08–2017–0186

Kovacevic, S., Nystuen, M., and Larsen, A. (2008). BIM integration with multi-lingual capabilities in construction projects. *Journal of Civil Engineering and Management*, 14(3), pp.247-254.

Kuhrmann, M., Kalus, G., and Then, M. (2014). The Process Enactment Tool Framework—Transformation of software process models to prepare enactment. *Science of Computer Programming*, *79*, pp.172–188.

Kumar, B. (2015). *A practical guide to adopting BIM in construction projects*. Whittles Publishing.

Kumar, B., (2024). *Notes on viva outcomes and feedback discussion*. [Notes from meeting held on 17 September 2024 regarding viva outcomes, University of Strathclyde]

Kumar, B., Cai, H., Hastak, M. (2017). An Assessment of Benefits of Using BIM on an Infrastructure Project. *International Conference on Sustainable Infrastructure 2017: Technology*.

Kumar, G., and Sharples, D. (2015). Design of Crossrail Farringdon Station: From an Engineer's Perspective. *ICE Publishing*. Available at: https://learninglegacy.crossrail.co.uk/documents/design-crossrail-farringdon-station-engineers-perspective-2/.

Kumar, M. (2024). *Building Information Modeling (BIM): Standards and best practices*. Wiley.

Kutner, M. H., Nachtsheim, C. J., Neter, J., and Li, W. (2005). *Applied linear statistical models*. McGraw-Hill.

Lai, H., Deng, X., Chen, H., Shi, P. (2020). Implementation of building information modelling standards for the Shanghai Metro, China. *Proceedings of the Institution of Civil Engineers-Municipal Engineer*. https://doi.org/10.1680/jmuen.18.00040.

Lam, K. P. (2004). Building Performance Simulation In The Singapore Construction Industry It Network. *Journal of Architectural and Planning Research*, *21*(4), 312–320.

Lancaster, F. D., Tobin, J. (2010). Integrated Project Delivery: Next-Generation BIM for Structural Engineering, in: Structures Congress 2010. Presented at the Structures Congress 2010, *American Society of Civil Engineers*, Orlando, Florida, United States, pp. 2809–2818. https://doi.org/10.1061/41130(369)254

Lang, S., and Baehr, C. (2012). Data Mining: A Hybrid Methodology for Complex and Dynamic Research. *College Composition and Communication*, *64*(1), 172–194.

Latiffi, A. A., Mohd, S., and Rakiman, U. S., 2016. Potential improvement of building information modelling (BIM) implementation in malaysian construction projects. In *Product Lifecycle Management in the Era of Internet of Things: 12th IFIP WG 5.1 International Conference, PLM 2015, Doha, Qatar, October 19–21, 2015, Revised Selected Papers 12* (pp. 149–158). Cham: Springer International Publishing.

Lawson, B., 2006. Embodied energy of building materials. *Environment Design Guide*, pp.1- 5.

Lee, S., Yu, J. (2016). Comparative Study of BIM Acceptance between Korea and the United States. J. *Constr. Eng. Manage*. 142, 05015016. https://doi.org/10.1061/(ASCE)CO.1943–7862.0001076

Lee, S., Yu, J., Jeong, D. (2015). BIM Acceptance Model in Construction Organizations. *J. Manage. Eng*. 31, 04014048. https://doi.org/10.1061/(ASCE)ME.1943–5479.0000252

Li, H., Ota, K., and Dong, M. (2018). Learning IoT in edge: Deep learning for the Internet of Things with edge computing. IEEE network, *32*(1), 96–101.

Li, J., Huang, X., and Wu, D. (2018). Data science and its application to the construction industry. *Construction Innovation*, 18(2), 133-149.

Li, K., Rollins, J., and Yan, E. (2017). Web of Science use in published research and review papers 199W2017: A selective, dynamic, cross-domain, content-based analysis. *Scientometrics*, *115*, 1–20.

Li, K.-H., and Georgiou, G. (2017). Systems architecture models: Crossrail design and delivery from the client's perspective. *Crossrail Ltd*. Available at: https://learninglegacy.crossrail.co.uk/documents/systems-architecture-models-crossrail-design-delivery-clients-perspective/.

Li, M. (2019). Passive Low Energy Consumption Building Cost Control Method in Coastal Green Cities. *Journal of Coastal Research*, *348*–353.

Li, N., and Wu, D. D. (2010). Using text mining and sentiment analysis for online forums hotspot detection and forecast. *Decision Support Systems*, *48*(2), 354–368.

Li, P., Zheng, S., Si, H., and Xu, K. (2019). Critical challenges for BIM adoption in small and medium-sized enterprises: Evidence from China. *Advances in Civil Engineering, 2019*, pp.1–14.

Li, T. H. Y., Ng, S. T., and Skitmore, M. (2012). Conflict or consensus: An investigation of stakeholder concerns during the participation process of major infrastructure and construction projects in Hong Kong. *Habitat International 36*, 333–342. https://doi.org/10.1016/j.habitatint.2011.10.012

Li, X., Lu, W., Xue, F., Wu, L., Zhao, R., Lou, J., and Xu, J. (2022). Blockchain-enabled IoT- BIM Platform (BIBP) for supply chain management in modular construction. *Journal of Construction Engineering and Management*, *148*(2), Article 04021195. https://doi.org/10.1061/(ASCE)CO.1943-7862.0002229

Liang, H., Zhang, S., and Su, Y. (2018). The structure and emerging trends of construction safety management research: A bibliometric review. *International Journal of Occupational Safety and Ergonomics*, *26*, 469–488.

Liang, V., Bock, C., and Zha, X. (2008). An ontological modelling platform. NIST.

Lill, I., and Witt, E. (2019.) 10th Nordic Conference on Construction Economics and Organization (Emerald Reach Proceedings Series, Vol. 2), *Emerald Publishing Limited*, *Bingley*, pp. 297–303.

Liu, A. (2019). Research on Fast Stretching Algorithm for Linear BIM Model of High-speed Railways. *China Mechanical Engineering* 30, 295–299. https://doi.org/10.3969/j.issn.1004–132X.2019.03.007

Liu, B. (2012). *Sentiment analysis and opinion mining*. Synthesis Lectures on Human Language Technologies, 5(1), 1-167. https://doi.org/10.2200/S00416ED1V01Y201204HLT016

Liu, B. (2017). Research on cloud rendering based 3D BIM model visualization techniques. *Journal of Beijing Jiaotong University* 41, 107–113. https://doi.org/10.11860/j.issn.1673–0291.2017.06.017

Liu, D., Chen, J., Hu, D., Zhang, Z. (2019). Dynamic BIM-augmented UAV safety inspection for water diversion project. *Computers in Industry*. https://doi.org/10.1016/j.compind.2019.03.004

Liu, H., Gao, X., and Wang, M. (2019). Integration of BIM and blockchain technology for effective facility management. *Automation in Construction*, 101, pp.60-69.

Liu, H., Cheng, J.C.P., Gan, V.J.L. and Zhou, S. (2022). A novel data-driven framework based on BIM and knowledge graph for automatic model auditing and quantity take-off. *Automation in Construction*, 139, p.101757. Available at: https://doi.org/10.1016/j.aei.2022.101757

Liu, Q., McCarthy, D. and Korhonen, A., (2020). Towards Better Context-aware Lexical Semantics: Adjusting Contextualized Representations through Static Anchors. *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Online, pp.4066–4075. Available at: https://aclanthology.org/2020.emnlp-main.333

Liu, S., Su, Y., Yin, H., Zhang, D., He, J., Huang, H., Jiang, X., Wang, X., Gong, H., Li, Z. and Xiu, H. (2021). An infrastructure with user-centered presentation data model for integrated management of materials data and services. *Npj Computational Materials*, 7(1), p.88.

Liu, W. (2019). Practice and Thoughts on Application of BIM Technology to Digital Railway Construction. *Journal of the China Railway Society* 41, 97–101. https://doi.org/10.3969/j.issn.1001–8360.2019.03.013

Liu, W., Guo, H., Li, H., Li, Y. (2015). Using BIM to Improve the Design and Construction of Bridge Projects: A Case Study of a Long-span Steel-box Arch Bridge Project (Retraction of vol 11, 125, 2014). *International Journal of Advanced Robotic Systems*. https://doi.org/10.5772/61239

Liu, X. and Issa, R.R.A., (2016). Unstructured data integration in construction: Challenges and opportunities. *Automation in Construction*, 65, pp.1-12.

Liu, X., Hu, W. (2019). Attention and sentiment of Chinese public toward green buildings based on Sina Weibo. *Sustainable Cities and Society 44*, 550–558. https://doi.org/10.1016/j.scs.2018.10.047

Liu, Y., Hur, J., Chan, W. K. B., Wang, Z., Xie, J., Sun, D., Handelman, S., Sexton, J., Yu, H., and He, Y. (2021). Ontological modelling and analysis of experimentally or clinically verified drugs against coronavirus infection. *Scientific data*, *8*(1), Article 16. https://doi.org/10.1038/s41597-021-00799-w

Liu, Y., van Nederveen, S., Hertogh, M. (2017). Understanding effects of BIM on collaborative design and construction: An empirical study in China. *International Journal of Project Management* 35, 686–698. https://doi.org/10.1016/j.ijproman.2016.06.007

Loosemore, M., (2015). The Role of Stakeholders in Project Success: Managing the Relationship between Contractors and Clients. *Construction Management and Economics*, 33(1), pp.67-75.

Lopez-Lira, A. (2023). Risk factors that matter: Textual analysis of risk disclosures for the cross-section of returns. *Jacobs Levy Equity Management Centre for Quantitative Financial Research Pape*r.

Love, P. E. D., and Matthews, J. (2019). The 'how' of benefits management for digital technology: From engineering to asset management. *Automation in Construction 107*, 102930. https://doi.org/10.1016/j.autcon.2019.102930

Love, P.E.D., Edwards, D.J., Smith, J., and Walker, D.H.T., (2020). The challenges of integrating big data and BIM: Overcoming obstacles in construction management. *Journal of Construction Engineering and Management*, 146(3), pp.1-11.

Lozano, S., Calzada-Infante, L., Adenso-Díaz, B., and Carbajal, S. G. (2019). Complex network analysis of keywords co-occurrence in the recent efficiency analysis literature. S*cientometrics*, *1*–21.

Lu, W., O'Connor, J.T. and Yang, K., (2020). Big Data and Construction: Opportunities and Challenges. *Journal of Construction Engineering and Management*, 146(1), p.04019113.

Luque, A., Mezher, T., and Hawi, M. (2019). Comparative analysis of BIM integration with data-driven decision-making in construction. *Computers in Industry*, 109, pp.100-114.

Luth, G. P., Schorer, A., Turkan, Y. (2014). Lessons from Using BIM to Increase Design-Construction Integration. Pract. *Period. Struct. Des. Constr*. 19, 103–110. https://doi.org/10.1061/(ASCE)SC.1943–5576.0000200

Ma, L. (2019). Application of Building Information Model (BIM) in the Design of Marine Architectural Structures. *Journal of Coastal Research*, *94*(SI), 117–120

Ma, Z., Zhang, D., Li, J. (2018). A dedicated collaboration platform for Integrated Project Delivery. *Automation in Construction 86*, 199–209. https://doi.org/10.1016/j.autcon.2017.10.024

MacDonald, I. (2016). Information handover principles. *Crossrail Ltd*. Available at: https://learninglegacy.crossrail.co.uk/documents/information-handover-principles/.

Mackenzie, N., and Knipe, S. (2006) *Research Dilemmas: Paradigms, Methods, and Methodology, Issues in Educational Research*, 16(2), pp. 193-205.

Mahamadu, A.-M., Mahdjoubi, L., Booth, C. A. (2013). Challenges to digital collaborative exchange for sustainable project delivery through building information modelling technologies. *Presented at the Sustainable City* 2013, Putrajaya, Malaysia, pp. 547–557. https://doi.org/10.2495/SC130461

Major Projects Authority (2016) GOV.UK. Available at: *https://www.gov.uk/government/groups/major-projects-authority#:~:text=The%20Major%20Projects%20Authority%20(MPA,we%20manage%20and%20deliver%20projects*. (Accessed: 12 August 2023).

Maksudovna, K.Z., Mirolyubova, A.N., and Yakusheva, A. (2023). Cost management in construction projects using BIM-based models. *Procedia Engineering*, 210, pp.17-24.

Malcolm, T. (2017). Crossrail: Building a Virtual Version of London’s Elizabeth line - The Development of the BIM Environment. *Crossrail Ltd*. Available at: https://learninglegacy.crossrail.co.uk/documents/crossrail-development-bim-environment/.

Manata, B., Miller, V. D., Mollaoglu, S., and Garcia, A. J. (2022). Documenting the Interactive Effects of Project Manager and Team-Level Communication Behaviors in Integrated Project Delivery Teams. *Project Management Journal*, *53*(1), 33–48.

Maqsoom, A., Rashid, H. B., Aslam, B., Ashraf, H., Abid, M., and Ejaz, A., 2020. BUILDING INFORMATION MODELLING APPLICATION FOR GROUNDWATER RECHARGE: DEVELOPMENT OF MULTIPLE STRUCTURES. *Environmental Engineering and Management Journal (EEMJ)*, 19(6).

Mariam, S., Khawaja, K. F., Qaisar, M. N., and Ahmad, F. (2022). Knowledge-oriented leadership, team cohesion, and project success: A conditional mechanism. *Project Management Journal*, *53*(2), 128–145.

Martin, R. (2017). Sustainability and Building Information Modelling (BIM). Environment Design Guide, *1*–18.

Martínez-Aires, M. D., López-Alonso, M., Martínez-Rojas, M. (2018). Building information modelling and safety management: A systematic review. *Safety Science 101*, 11–18. https://doi.org/10.1016/j.ssci.2017.08.015

Marzouk, M. and Enaba, M. (2019). Applying text mining in AEC industry to enhance accident report analysis. *Automation in Construction*, 98, pp.275-288.

Marzouk, M., Enaba, M. (2019). Text analytics to analyze and monitor construction project contract and correspondence. *Automation in Construction 98*, 265–274. https://doi.org/10.1016/j.autcon.2018.11.018

May, A. I., Taylor, M., and Irwin, D. (2017). Crossrail: A Case Study in BIM.

Mayouf, M., Gerges, M., and Cox, S. (2019). 5D BIM: An investigation into the integration of quantity surveyors within the BIM process. *Journal of Engineering, Design and Technology*, *17*(3), 537–553.

Mazairac, A. R., and Beetz, J. (2013). A data-driven approach to structural design optimization in BIM. *Automation in Construction*, 31, 70-82.

McCarthy, J. (2016). Lindsey Street Bridge: A Structural Solution to Settlement Mitigation. *Crossrail Ltd*. Available at: https://learninglegacy.crossrail.co.uk/documents/lindsey-street-bridge-structural-solution-settlement-mitigation/.

McCarthy, J., Hayes, P. J. (1981). Some Philosophical Problems from the Standpoint of Artificial Intelligence, in: *Readings in Artificial Intelligence*. Elsevier, pp. 431–450. https://doi.org/10.1016/B978–0-934613–03–3.50033–7.

McFadden, E.M. and Ernzen, J.J., 2003. Partnered project performance at the City of Phoenix. In *Construction Research Congress: Wind of Change: Integration and Innovation* (pp. 1-9).

McKinsey and Company. (2020) *The state of AI in 2020*. Available at: https://www.mckinsey.com (Accessed: October 2024).

Meller, T. (2013) Research methodologies for IT projects. *International Journal of IT Research*, 7(4), pp.120-134.

Menches, C. L., Hanna, A. S. (2006). Quantitative Measurement of Successful Performance from the Project Manager's Perspective. *J. Constr. Eng. Manage*. 132, 1284–1293. https://doi.org/10.1061/(ASCE)0733–9364(2006)132:12(1284)

Merschbrock, C. and Munkvold, B.E., (2015). Information Technology in the AEC Industry: Challenges and Opportunities. *Automation in Construction*, 58, pp. 1-8.

Methods for Analyzing Unstructured Information.

Meyer, K. P. (2016). Technology in Architectural Practice: Transforming Work with Information, *1960*s-1990s. *Information and Culture, 51*(2), 249–266.

Miceli, R., Di Paola, M. and Sannino, A., (2019). BIM: A tool for collaborative design and building lifecycle management. *Automation in Construction*, 103, pp.60-70. Available at: https://doi.org/10.1016/j.autcon.2019.01.004

Mike, M. (2018). *BREEAM for Underground Stations*. Crossrail Ltd. Available at: https://learninglegacy.crossrail.co.uk/documents/breeam-for-underground-stations/.

Mill, J.S. (1843) *A System of Logic, Ratiocinative and Inductive: Being a Connected View of the Principles of Evidence, and the Methods of Scientific Investigation.* London: John W. Parker.

Miner, G. (Ed.). (2012). Practical text mining and statistical analysis for non-structured text data applications, *1*st ed. ed. *Academic Press*, Waltham, MA.

Mitesh, K., and Aliki, P. (2016). Design of the Deep Cut and Cover Crossrail Paddington Station Using Finite Element Method. *Crossrail Ltd*. Available at: https://learninglegacy.crossrail.co.uk/documents/design-deep-cut-cover-crossrail-paddington-station-using-finite-element-method/.

Moayedi, H., Moayedi, J., and Hedayati, H. (2019). Use of BIM in geotechnical and tunneling projects. *Tunnelling and Underground Space Technology*, 87, pp.122-130.

Mok, K. Y., Shen, G. Q., Yang, J. (2015). Stakeholder management studies in mega construction projects: A review and future directions. *International Journal of Project Management 33*, 446–457. https://doi.org/10.1016/j.ijproman.2014.08.007

Mok, M. K. Y., and Shen, G. Q. (2016). A network-theory Based Model for Stakeholder Analysis in Major Construction Projects. *Procedia Engineering 164*, 292–298. https://doi.org/10.1016/j.proeng.2016.11.622

Moon, H., Dawood, N., and Kang, L. (2014a). Development of workspace conflict visualization system using 4D object of work schedule. *Advanced Engineering Informatics*. https://doi.org/10.1016/j.aei.2013.12.001

Moon, H., Kim, H., Kim, C., and Kang, L. (2014b). Development of a schedule-workspace interference management system simultaneously considering the overlap level of parallel schedules and workspaces. *Automation in construction*. https://doi.org/10.1016/j.autcon.2013.06.001

Moon, H., Park, C., and Kim, H. (2018). Text mining-based design information extraction for smart city infrastructure development. *Automation in Construction*, 94, pp.1-8.

Mordue, S. (2016) *Introducing the periodic table of BIM, NBS*. Available at: https://www.thenbs.com/knowledge/introducing-the-periodic-table-of-bim (Accessed: 07 December 2023).

Moreno, A., and Redondo, T. (2016). Text analytics: The convergence of Big Data and artificial intelligence. IJIMAI, *3*(6), 57–64.

Moreno, C., Olbina, S., and Issa, R. R. (2019). BIM use by architecture, engineering, and construction (AEC) industry in educational facility projects. Advances in Civil Engineering, 2019.

Morgan, D.L. (2007) Paradigms lost and pragmatism regained: Methodological implications of combining qualitative and quantitative methods. *Journal of Mixed Methods Research,* 1(1), pp. 48-76.

Morgan, D.L. (2014) *Pragmatism as a Paradigm for Social Research. Qualitative Inquiry*, 20(8), pp. 1045-1053.

Morgan, H., and Fung, W. (2018). Performance Benefits Associated with Elevated Temperature Set Points and Increased Error Deadbands for Room Temperature Controllers. *Crossrail Ltd*, ICE Publishing. Available at: https://learninglegacy.crossrail.co.uk/documents/performance-benefits-associated-elevated-temperature-set-points-increased-error-deadbands-room-temperature-controllers/.

Morgans, N. (2021). Final Design Overview: A Crossrail Design Assurance Process. *Crossrail Ltd*, ICE Publishing. Available at: https://learninglegacy.crossrail.co.uk/documents/final-design-overview-a-crossrail-design-assurance-process/.

Moss, S., and Are, F. (2003). Challenges in data mining projects. *Journal of Data Mining and Knowledge Discovery*, 7(4), 401-425.

Motwani, J., Kumar, A., and Novakoski, M. (1995). Measuring construction productivity: A practical approach. *Work Study*, *44*(8), 18–20.

Moxton, S., and Atherton, S. (2014). The importance of construction mock-ups and trials. *Crossrail Ltd*. Available at: https://learninglegacy.crossrail.co.uk/documents/the-importance-of-construction-mock-ups-and-trials/.

Mtya, A., and Windapo, A. O. (2019). Drivers and Barriers to the Adoption of Building Information Modelling (BIM) By Construction Firms in South Africa. *Innovative Production and Construction*.

Naamane, A., Boukara, A. (2015). A Brief Introduction to Building Information Modelling (BIM) and its interoperability with TRNSYS. *Renewable Energy and Sustainable Development*.

Nadali, A., Kakhky, E. N., and Nosratabadi, H. E. (2011), April. Evaluating the success level of data mining projects based on CRISP-DM methodology by a Fuzzy expert system. In 2011 3rd International Conference on Electronics Computer Technology (Vol. 6, 161–165). IEEE.

Naeem, M., Jamal, T., Diaz-Martinez, J., Butt, S. A., Montesano, N., Tariq, M. I., De-la-Hoz-Franco, E., and De-La-Hoz-Valdiris, E. (2022). Trends and future perspective challenges in big data. In Advances in Intelligent Data Analysis and Applications: Proceeding of the Sixth Euro-China Conference on Intelligent Data Analysis and Applications, *15*–18 October 2019, Arad, Romania (pp. 309–325). Singapore: Springer Singapore.

Naeem, S., Usman, M., and Khan, S. (2022). Big Data in construction management. *Construction Innovation*, 22(1), 79-97.

Nasila, M., and Cloete, C., 2018. Adoption of Building Information Modelling in the construction industry in Kenya. *Acta Structilia*, *25*(2), 1–38.

Nasir, H., Ahmed, H., Haas, C., Goodrum, P. M. (2014). An analysis of construction productivity differences between Canada and the United States. *Construction*

Management and Economics 32, 595–607. https://doi.org/10.1080/01446193.2013.848995

Natarajan, M. (2005). Role of text mining in information extraction and information management. *DESIDOC Journal of Library and Information Technology*, *25*(4).

National Audit Office [NAO]. (2013). Over-optimism in government projects. London: National Audit Office.

National BIM Standard Project Committee. (2007). National building information modelling standard, version 1, Part 1: Overview, principles, and methodologies. *National Institute of Building Sciences*.

National Building Specification [NBS] (2012–2019). NBS National BIM Report 2012. Newcastle Upon Tyne: NBS.

Nauck, D., and Kruse, R. (2019). *Foundations of fuzzy systems*. Cham: Springer.

Nawaz, A., Su, X., and Nasir, I. (2021). BIM Adoption and its impact on planning and scheduling influencing mega plan projects - (CPEC-) quantitative approach. *Complexity*, vol. 21, 9 pages.

Nayak, A. S., Kanive, A. P., Chandavekar, N., and Balasubramani, R. (2016). Survey on pre- processing techniques for text mining. *International Journal of Engineering and Computer Science*, *5*(6), 16875–16879.

NBS (2021). *Uniclass: A Unified Classification System for the Construction Industry*. National Building Specification (NBS), NBS Enterprises Ltd, UK. Available at: https://www.thenbs.com/our-tools/uniclass-2015 (Accessed: October 2024).

Neuman, W.L. (2009) *Social Research Methods: Qualitative and Quantitative Approaches*. 7th ed. Boston: Allyn and Bacon.

Ng, H. S., Toukourou, A., Soibelman, L. (2006). Knowledge Discovery in a Facility Condition Assessment Database Using Text Clustering. *J. Infrastruct. Syst*. 12, 50–59. https://doi.org/10.1061/(ASCE)1076–0342(2006)12:1(50)

Ning, X., Zhu, Q., Ren, X., Han, Z., Wang, H., Zhao, W., Zhang, H. (2018). Graphical Processing Method of High-Speed Railway Construction Progress Data. *Journal of Southwest Jiaotong University* 53, 414–419.

Nowotny, H., Scott, P., and Gibbons, M. (2003*) 'Mode 2 Revisited: The New Production of Knowledge'*, Minerva, 41, pp.179-194.

O'Connor, J.T. and Yang, K., (2004). Managing Data Complexity in Construction Projects. *Journal of Computing in Civil Engineering*, 18(4), pp. 327-334.

Oesterreich, T. D., and Teuteberg, F. (2016.) Understanding the implications of digitisation and automation in the context of Industry 4.0: A triangulation approach and elements of a research agenda for the construction industry. *Computers in industry*, *83*, pp.121–139.

Okoli, C. (2023) *Introduction to Research Methods in Social Sciences*. 2nd edn. London: Routledge.

Oktem, S., Akcamete, A., and Ergen, E. (2020). BIM implementation in infrastructure projects. In E. Ademci and S. Gundes (Eds.), *Review of studies on BIM adoption in AEC industry, 5th International Project and Construction Management Conference (IPCMC) Proceedings*. Istanbul Technical University.

Olanrewaju, O. I., Chileshe, N., Babarinde, S. A., and Sandanayake, M. (2020). Investigating the barriers to building information modelling (BIM) implementation within the Nigerian construction industry. *Engineering, Construction and Architectural Management*, *27*, 2931–2958.

Olanrewaju, O. I.; Kineber, A. F.; Chileshe, N.; and Edwards, D.J. (2021). Modelling the Impact of Building Information Modelling (BIM) Implementation Drivers and Awareness on Project Lifecycle. *Sustainability, 13*, 8887.

Olawumi, T. O., Chan, D. W. M., and Wong, J. S. (2017). Evolution in the intellectual structure of BIM research: A bibliometric analysis. *Journal of Civil Engineering and Management*, *23,* 1060–1081.

Oraee, M., Hosseini, M. R., Papadonikolaki, E., Palliyaguru, R., Arashpour, M. (2017). Collaboration in BIM-based construction networks: A bibliometric-qualitative literature review. *International Journal of Project Management 35*, 1288–1301. https://doi.org/10.1016/j.ijproman.2017.07.001

Oshodi, O. S., Awuzie, B. O., Akotia, J., Ademiloye, A. S., and Ngowi, A. (2020). A bibliometric analysis of recycled concrete research. (1978–2019). *Built Environment Project and Asset Management*, *10(5)*, 725–736.

Osman, A., Ling, L.H., and Sulaiman, M.A. (2018) ‘Application of Deductive and Inductive Approaches in Research’, *Journal of Educational Research and Practice*, 8(4), pp. 72-81.

Othman, I. and Nadim, S., (2010). Stakeholder Engagement in Construction Projects: A Review. International Journal of Project Management, 28(6), pp. 508-515.

Oti-Sarpong, K., Pärn, E. A., Burgess, G., and Zaki, M. (2022). Transforming the construction sector: An institutional complexity perspective. *Construction innovation*, *22*(2), 361- 387.

Pan, B., Luo, T., and Zhang, X. (2020). A patent analysis-based approach to BIM for construction industry innovation. *Automation in Construction,* 109, p.102987.

Pan, X., Zhong, B., Wang, X., and Xiang, R. (2021). Text mining-based patent analysis of BIM application in construction. *Journal of Civil Engineering and Management*, *27*(5), 303–315.

Papadonikolaki, E. (2018). Loosely coupled systems of innovation: Aligning BIM adoption with implementation in Dutch construction. *Journal of management in engineering*, *34*(6), p.05018009.

Park, H., Lee, S., and Jang, D. (2023) Principles of Positivism in Modern Social Science Research, *Journal of Applied Research Methods,* 12(2), pp. 89-101.

Park, H.-S., Thomas, S. R., Tucker, R. L. (2005). Benchmarking of Construction Productivity. *J. Constr. Eng. Manage*. 131, 772–778. https://doi.org/10.1061/(ASCE)0733–9364(2005)131:7(772)

Parker, N., Bannister, P., Jackson, Q., and Stoller, P. (2017). Optimising environmental performance using building performance simulation. *Environment Design Guide*, *1*– 16.

Pärn, E. A.; Edwards, D. J.; and Sing, M.C. (2017). The building information modelling trajectory in facilities management: A review. *Autom. Constr*, *75*, 45–55.

Parvaiz, G.S., Mufti, O., and Wahab, M. (2016) 'Pragmatism in Research: A Review of Key Paradigms', *International Journal of Quantitative and Qualitative Research Methods*, 4(1), pp. 10-17.

Patanakul, P., Kwak, Y. H., Zwikael, O., Liu, M. (2016). What impacts the performance of large-scale government projects? *International Journal of Project Management* 34, 452–466. https://doi.org/10.1016/j.ijproman.2015.12.001.

Pauwels, P., and Zhang, X. (2015). BIM-based data modeling: Opportunities and challenges. *Construction Research Congress*, 174-183.

Pekuri, A., Haapasalo, H., and Herrala, M. (2011). Productivity and performance management–managerial practices in the construction industry. *International Journal of Performance Measurement*, *1*(1), 39–58.

Peng, J. (2017). The application of BIM technology in simulation modelling of concrete dam construction. *Agro Food Industry Hi-Tech* 28, 1757–1761.

Pettinella, M. (2020). A Connected Future: Digital Transformation in European Construction Firms and Beyond. [Blog] *Autodesk Construction Cloud*. https://constructionblog.autodesk.com/european-construction/

Phoenix, C., Osborne, N.J., Redshaw, C., Moran, R., Stahl-Timmins, W., and Depledge, M.H. (2013) '*Paradigmatic approaches to studying environment and human health interactions: Evidence, philosophy, and methods', Environmental Science and Policy*, 25, pp. 218-228.

Piatetsky, G. (2022). *The CRISP-DM methodology: Past, present, and future*. Data Mining Book, pp. 13-25.

Poirier, E. A., Staub-French, S., and Forgues, D. (2015). Measuring the impact of BIM on labor productivity in a small specialty contracting enterprise through action-research. *Automation in construction*, *58*, pp.74–84. https://doi.org/10.1016/j.autcon.2015.07.002

Porwal, A., and Hewage, K. N. (2013). Building Information Modelling (BIM) partnering framework for public construction projects. *Automation in construction*, *31*, pp.204–214.

Pour Rahimian, F., Seyedzadeh, S., Oliver, S., Rodriguez, S., Dawood, N. (2020). On-demand monitoring of construction projects through a game-like hybrid application of BIM and machine learning. *Automation in Construction 110*, 103012. https://doi.org/10.1016/j.autcon.2019.103012

Power, D. J., and Heavin, C. (2018). Data-Based Decision Making and Digital Transformation: Nine Laws for Success. *Business Expert Press*.

Pritchard, A. (1969). Statistical bibliography or bibliometrics. *Journal of Documentation*, *25*, 348.

Pryke, S. (2012). Social network analysis in construction. London: John Wiley and Sons.

Pryke, S. D. (2004). Analysing construction project coalitions: Exploring the application of social network analysis. *Construction Management and Economics* 22, 787–797. https://doi.org/10.1080/0144619042000206533

Pujara, J., Miao, H., Getoor, L. and Cohen, W. (2013). Knowledge graph identification. In *The Semantic Web–ISWC 2013: 12th International Semantic Web Conference, Sydney, NSW, Australia, October 21-25, 2013, Proceedings, Part I 12* (pp. 542-557). Springer Berlin Heidelberg.

Qu, Y. (2021). BIM integration in tunnel construction for project efficiency. *Journal of Infrastructure Systems*, 27(4), pp.04021041.

Queensland Government. (2021). Building information modelling (BIM) for transport and main roads. A guide to enabling BIM on road infrastructure projects.

Rahim, N. S. A. and Zakaria, S. A. S. (2017). The framework of building information modelling (BIM) implementation stages for Malaysian construction industry. *IPTEK Journal of Proceedings Series*, *3*(6).

Rajman, M., and Besançon, R. (1998). Text mining: A new paradigm in text analysis. *International Journal of Data Mining and Knowledge Discovery*, 2(1), 1-14.

Rajput, A. (2020). *Challenges in text mining and natural language processing*. In Proceedings of the 2020 International Conference on Machine Learning and Data Engineering (pp. 98-102). IEEE. https://doi.org/10.1109/ICMLDE50209.2020.9242955

Rajput, A. (2020). Natural language processing, sentiment analysis, and clinical analytics. In Innovation in health informatics (pp. 79–97). Academic Press.

Raouf, A.M, and Al-Ghamdi, S.G. (2019). Building information modelling and green buildings: Challenges and opportunities. *Archit. Eng. Des. Manag*, *15*, 1–28.

Rehurek, R., and Sojka, P. (2011). Gensim–python framework for vector space modelling. NLP Centre, Faculty of Informatics, *Masaryk University, Brno, Czech Republic*, *3*(2), p.2.

Rendell, D., and Thurlow, P. (2014). Data management, analysis and visualisation on Crossrail Drive X Western Tunnels. *Crossrail Ltd*. Available at: https://learninglegacy.crossrail.co.uk/documents/data-management-analysis-and-visualisation-on-crossrail-drive-x-western-tunnels/.

Retzlaff, T. (2018). Level of BIM implementation in the field of infrastructure development. *Bautechnik*, *95*(7), 512–517. https://doi.org/10.1002/bate.201800045

Rivas, R. A., Borcherding, J. D., González, V., Alarcón, L. F. (2011). Analysis of Factors Influencing Productivity Using Craftsmen Questionnaires: Case Study in a Chilean Construction Company. *J. Constr. Eng. Manage*. 137, 312–320. https://doi.org/10.1061/(ASCE)CO.1943–7862.0000274

Rojas, E. M., Aramvareekul, P. (2003). Labor Productivity Drivers and Opportunities in the Construction Industry. *J. Manage. Eng*. 19, 78–82. https://doi.org/10.1061/(ASCE)0742–597X(2003)19:2(78)

RStudio Team. (2020). RStudio: Integrated Development for R. RStudio, PBC, Boston, MA URL http://www.rstudio.com/.

Ruiz, J., Martinez, J., and Crespo, R. (2021). The application of artificial intelligence in project management research: A review. *International Journal of Interactive Multimedia and Artificial Intelligence*, *6*(6), 1–13. http://dx.doi.org/10.9781/ijimai.2020.12.003

Russell, S., and Norvig, P. (2010). *Artificial Intelligence: A modern approach* (3rd ed.). London: Prentice Hall.

Ryan, G. (2018) '*Introduction to Positivism, Interpretivism and Critical Theory*', Nurse Researcher, 25(4), pp. 41-49.

Ryburn, J., Mann, H., Doran, S., and Cheung, K. (2014). Design of Crossrail's precast tunnel linings for fire. *Crossrail Ltd*. Available at: https://learninglegacy.crossrail.co.uk/documents/design-of-crossrails-precast-tunnel-linings-for-fire/.

Sacks, R., Barak, R. (2008). Impact of three-dimensional parametric modelling of buildings on productivity in structural engineering practice. *Automation in Construction* 17, 439–449. https://doi.org/10.1016/j.autcon.2007.08.003

Sacks, R., Gurevich, U., and Shrestha, P. (2016). A review of building information modelling protocols, guides and standards for large construction clients. *Journal of Information Technology in Construction* (ITcon), 21(29), 479–503.

Sajovic, I., Tomc, H. G., and Podgornik, B. (2018). Bibliometric study and mapping of a journal in the field of visualization and computer graphics. *COLLNET Journal of Scientometrics and Information Management*, *12*, 263–287.

Saka and Chan. (2019). A scientometric review and metasynthesis of building information modelling (BIM) research in Africa. Buildings, *9*(4), p.85.

Saka, A. B., and Chan, D. W. M. (2020). Profound barriers to building information modelling (BIM) adoption in construction small and medium-sized enterprises (SMEs). *Construction Innovation*: Information, Process, Management, *20*, 261–284.

Salehi, S. A., and Yitmen, I. (2018). Modelling and analysis of the impact of BIM-based field data capturing technologies on automated construction progress monitoring. *International Journal of Civil Engineering*, *16*, 1669–1685.

Saltz, R. M. (2015). *Text mining and its applications in construction management*. Wiley.

Sampaio, R. B., Fonseca, M. V. D. A., and Zicker, F. (2016). Co-authorship network analysis in health research: Method and potential use. *Health research policy and systems*, *14*(1), 1–10.

Sankaran, B., Nevett, G., O'Brien, W. J., Goodrum, P. M., Johnson, J. (2018). Civil Integrated Management: Empirical study of digital practices in highway project delivery and asset management. *Automation in Construction* 87, 84–95. https://doi.org/10.1016/j.autcon.2017.12.006.

Santos, R., Costa, A. A., and Grilo, A. (2017). Bibliometric analysis and review of Building Information Modelling literature published between 2005 and 2015. *Automation in Construction*, *80*, pp.118–136.

Santoso, D.S. and Gallage, P.G.M.P., 2020. Critical factors affecting the performance of large construction projects in developing countries: A case study of Sri Lanka. *Journal of Engineering, Design and Technology*, *18*(3), pp.531-556.

Saunders, M., Lewis, P., and Thornhill, A. (2016) *Research Methods for Business Students*. 7th ed. Harlow: Pearson Education.

Saunders, M., Lewis, P., and Thornhill, A. (2019) *Research Methods for Business Students.* 8th ed. Harlow: Pearson Education.

Scherer, R. and Reul, M. (2000). Early use of TM in the AEC sectors. Journal of Information Technology in Construction, 5, pp.1-15.

Schölkopf, B., and Smola, A. J. (2002). Learning with kernels: Support vector machines, regularization, optimization, and beyond. MIT press.

Schwade, F., (2021). Social Collaboration Analytics Framework: A framework for providing business intelligence on collaboration in the digital workplace. Decision Support Systems, *148*, Article 113587.

Schwartzkopf, W. (1995). Calculating Lost Labor Productivity in Construction Claims.

Schwartzkopf, W. (2004). Calculating lost labor productivity in construction claims, *2*nd ed. ed. *Aspen Publishers*, New York, NY.

Scotland, J. (2012) '*Exploring the philosophical underpinnings of research: Relating ontology and epistemology to the methodology and methods of the scientific, interpretive, and critical research paradigms', English Language Teaching*, 5(9), pp. 9-16.

Scott, J. (2005). Social Network Analysis as an Analytic Tool for Interaction Patterns in Primary Care Practices. *The Annals of Family Medicine* 3, 443–448. https://doi.org/10.1370/afm.344

Sebastian, R. (2010, May). Breaking through business and legal barriers of open collaborative processes based on building information modelling (BIM). In CIM World Building Congress (pp. 166–186). sector: An institutional complexity perspective. *Construction Innovation*.

Segal, T. (2022). Understanding a Decision Support System (DSS). *Investopedia*. https://www.investopedia.com/terms/d/decision-support-system.asp

Shahinmoghadam, M., and Motamedi, A. (2019). Review of BIM-centred IoT Deployment: State of the Art, Opportunities, and Challenges.

Shamim, M.M.I. (2024). *Artificial Intelligence in Project Management: Enhancing Efficiency and Decision-Making*. EDGE Bangladesh. DOI: 10.62304/ijmisds.v1i1.107.

Sheil, M. R. Thomsen, M. Tamke, and S. Hanna. (2020.), Design Transactions: Rethinking Information Modelling for a New Material Age (pp. 68–73). London: UCL Press.

Shen, L., Shen, E., Luo, Y., Yang, X., Hu, X., Zhang, X., Tai, Z., and Wang, J. (2022). Towards natural language interfaces for data visualization: A survey. *IEEE Transactions on Visualization and Computer Graphics*.

Shen, Q., Li, B., and Xue, F. (2022). Text mining and its applications in construction. *Automation in Construction*, 48, 52-66.

Shi, Z. (Michael), Lee, G. M., and Whinston, A. B. (2016). Toward a Better Measure of Business Proximity: Topic Modelling for Industry Intelligence. *MIS Quarterly*, *40*(4), 1035–1056.

Shin and Bulut (2021, April 8). Okan Bulut: Text Vectorization Using Python: Term-Document Matrix. *Retrieved from https://okan.cloud/posts/2021-04-08-text-vectorization-using-python-term-document-matrix/*

Shin, M. H., Lee, H. K., Kim, H. Y. (2018). Benefit-Cost Analysis of Building Information Modelling (BIM) in a Railway Site. *Sustainability*. https://doi.org/10.3390/su10114303

Shmueli, G., and Koppius, O. R. (2011). Predictive Analytics in Information Systems Research. *MIS Quarterly*, *35*(3), 553–572.

Shou, W., Wang, J., Wang, X., Chong, H. Y. (2015). A Comparative Review of Building Information Modelling Implementation in Building and Infrastructure Industries. *Arch Computat Methods Eng* 22, 291–308. https://doi.org/10.1007/s11831–014–9125–9

Sibert, B. (2013). Using building information modelling on a highway project. *Proceedings of the Institution of Civil Engineers: Civil Engineering 166*, 9. https://doi.org/10.1680/cien.2013.166.1.9

Siggelkow, N. (2007) Persuasion with case studies, *Academy of Management Journal*, 50(1), pp. 20-24. https://doi.org/10.5465/amj.2007.24160882.

Singh, V. (2019). Digitalization, BIM ecosystem, and the future of built environment: How widely are we exploring the different possibilities? *ECAM ahead-of-print*. https://doi.org/10.1108/ECAM-01–2018–0004

Skarpathiotaki, C. G., and Psannis, K. E. (2022). Cross-industry process standardization for text analytics. Big Data Research, *27*, p.100274.

Skarpathiotaki, M., and Psannis, K. (2022). Big Data analytics: Applications and challenges in data mining. *International Journal of Data Science*, 5(3), 179-191.

Smiraglia, R. P. (2015). Domain analysis of domain analysis for knowledge organization: Observations on an emergent methodological cluster. *KO Knowledge Organization*, 42(8), 602–614.

Smith, C. J.; and Wong, A. T.C. (2022). Advancements in artificial intelligence-based decision support systems for improving construction project sustainability: A systematic literature review. *In Informatics* (Vol. 9, No. 2, p. 43). MDPI.

Smith, J., (2020). Building Information Modeling (BIM) for Improved Project Performance: Communication, Coordination, and Collaboration. *Journal of Construction Management*, 34(2), pp. 142-155.

Smith, R. (2015). Examples of Specialist Grouting on Contract C310 Thames Tunnels. *Crossrail Ltd*. Available at: https://learninglegacy.crossrail.co.uk/documents/examples-specialist-grouting-contract-c310-thames-tunnels/.

Sobhkhiz, S., and El-Diraby, T. (2023). BIM and unstructured data: A gap analysis. *Journal of Construction Engineering and Management*, 149(1), 04022091.

Son, B.-Y., and Lee, E.-B. (2019). Using text mining to estimate schedule delay risk of 13 offshore oil and gas EPC case studies during the bidding process. *Energies*, *12*(10), p.1956.

Song, Z., Shi, G., Wang, J., Wei, H., Wang, T., Zhou, G. (2019). Research on management and application of tunnel engineering based on BIM technology. *Journal Of Civil Engineering and Management*. https://doi.org/10.3846/jcem.2019.11056

Ssenyonga, A. (2018). An Investigation of Surface Settlement and Volume Loss Associated with SCL Tunnelling at Stepney Green. *Crossrail Ltd*, ICE Publishing. Available at: https://learninglegacy.crossrail.co.uk/documents/an-investigation-of-surface-settlement-and-volume-loss-associated-with-scl-tunnelling-at-stepney-green/.

St. John, A., Weinmar, W., Fornelli, D., and McGirr, D. (2016). Design and construction of Fisher Street crossover cavern on Crossrail contract C300/C410, London. *ICE Publishing*. Available at: https://learninglegacy.crossrail.co.uk/documents/design-construction-fisher-street-crossover-cavern-crossrail-contract-c300c410-london/.

Stake, R.E. (1995) *The Art of Case Study Research*. Sage Publications, Thousand Oaks, CA.

Stakeholder Democracy Network. (2014). *Solar lanterns in Niger Delta communities: Experiences in building a sustainable distribution model*. International Institute for Environment and Development.

Stanisz, E., Kamiński, T., and Wojciechowski, T. (2019). BIM in tunnel construction: Challenges and best practices. *Tunnelling and Underground Space Technology*, 83, pp.96-104.

Stenlund, B. (2005). The Galaxen model—a concept for rehabilitation and prevention in the construction industry. *Scandinavian Journal of Work, Environment and Health*, pp.110- 115.

Stojanovska-Georgievska, L., Trifunovic, A., and Mitrevski, V. (2022). BIM integration with sustainability practices in construction. *Journal of Sustainable Development of Energy, Water and Environment Systems*, 10(2), pp.156-168.

Stolterman, E., and Fors, A. C. (2004). Information technology and the good life. *In Information systems research* (pp. 687–692). Boston, MA: Springer.

Stowell, G. (2018). Client Advice and Technical Services to Chief Engineer's Group. *Crossrail Ltd*, ICE Publishing. Available at: https://learninglegacy.crossrail.co.uk/documents/client-advice-and-technical-services-to-chief-engineers-group/.

Su, J., and Thomas, A. (2014). Design of Sprayed Concrete Linings in Soft Ground: A Crossrail Perspective. *Crossrail Ltd*. Available at: https://learninglegacy.crossrail.co.uk/documents/design-sprayed-concrete-linings-soft-ground%E2%94%80-crossrail-perspective/.

Succar, B. (2009). Building information modelling framework: A research and delivery foundation for industry stakeholders. *Automation in Construction 18*, 357–375. https://doi.org/10.1016/j.autcon.2008.10.003

Succar, B. (2013). Building Information Modelling: Conceptual constructs and performance improvement tools. *School of Architecture and Built Environment Faculty of Engineering and Built Environment, University of Newcastle*: Newcastle.

Suermann, P. C. (2009). *Evaluating the impact of building information modelling (BIM) on construction*. University of Florida.

Suermann, P. C., and Issa, R. R. (2009). Evaluating industry perceptions of building information modelling (BIM) impact on construction. *Journal of Information Technology in Construction* (ITcon), 14(37), 574–594.

Sun, M., Cao, Q., and Zhang, Z. (2018). Text data cleaning in text mining. *Journal of Computer Science and Technology*, 33(2), 281-298. https://doi.org/10.1007/s11390-018-1832-5

Sun, W., Cai, Z., Li, Y., Liu, F., Fang, S., and Wang, G. (2018). Data processing and text mining technologies on electronic medical records: A review. Journal of healthcare engineering, *2018*.

Tahir, A. (2017). *BIM Metrics*. Crossrail Ltd. Published 14 March 2017. Available at: https://learninglegacy.crossrail.co.uk/documents/bim-metrics/.

Tan, P. N. (1999). *Text mining: Products and applications*. Springer Science and Business Media.

Tang, L., Zhang, G., and Zhang, J. (2019). Text mining and BIM: A new frontier in construction management. *Journal of Management in Engineering*, 35(1), 04018060.

Tashakkori, A., and Teddlie, C. (2003) *Handbook of Mixed Methods in Social and Behavioral Research*. Thousand Oaks, CA: SAGE Publications.

Taylor, M. (2018). Crossrail Project: Application of BIM (Building Information Modelling) and Lessons Learned. *Crossrail Ltd*, ICE Publishing. Available at:

https://learninglegacy.crossrail.co.uk/documents/crossrail-project-application-of-bim-building-information-modelling-and-lessons-learned/.

Teddlie, C. and Tashakkori, A. (2010) *Mixed Methodology: Combining Qualitative and Quantitative Approaches*. 2nd edn. Thousand Oaks, CA: Sage Publications.

Teicholz, P. (Ed.). (2013). *BIM for facility managers*. London: John Wiley and Sons.

Teicholz, P. M. (2004). *Labor-productivity declines in the construction industry: Causes and remedies*. New York, NY: AECbytes Viewpoint.

Teicholz, P. M. (2013). *Labor-productivity declines in the construction industry: Causes and remedies (a second look)*. New York, NY: AECbytes Viewpoint.

Teng, X., Chen, L., and Wang, S. (2022). Enhanced BIM applications using text mining for lifecycle management. *Journal of Computing in Civil Engineering*, 36(4), pp.04022029.

Teng, Y., Xu, J., Pan, W., and Zhang, Y. (2022). A systematic review of the integration of building information modelling into life cycle assessment. *Building and Environment*, *221*, Article 109260. https://doi.org/10.1016/j.buildenv.2022.109260

Teo, M. M., and Loosemore, M. (2011). Community-based protest against construction projects: A case study of movement continuity. *Construction Management and Economics*, *29*, 131–144. https://doi.org/10.1080/01446193.2010.535545

Tereshko, V., and Loengarov, A. (2005). Collective decision-making in honey bee foraging dynamics. *Computing and Information Systems*, *9*(3), 1–7.

Thomas, H. R., and Završki, I. (1999). Construction baseline productivity: Theory and practice. *Journal of Construction Engineering and Management*, *125*, 295–303. https://doi.org/10.1061/(ASCE)0733–9364(1999)125:5(295).

Thomas, H. R., Maloney, W. F., Horner, R. M. W., Smith, G. R., Handa, V. K., and Sanders, S. R. (1990). Modelling construction labor productivity. *Journal of Construction Engineering and Management*, *116*, 705–726. https://doi.org/10.1061/(ASCE)0733–9364(1990)116:4(705).

Tookey, J. E. (2012). Shaving BIM: Establishing a framework for future BIM research in New Zealand. *International Journal of Construction Supply Chain Management*, *2*(2), 66–79. https://doi.org/10.14424/ijcscm201012-66-79

Törneman, C. (2015). *Collaboration in construction project delivery study* [Master's thesis, Chalmers University of Technology]. Chalmers Open Digital Repository. https://odr.chalmers.se/server/api/core/bitstreams/d22bb522-e992-43f2-ae74-0a7ed2e63570/content

Tran, V. D., and Tookey, J. E. (2011). Labour productivity in the New Zealand construction industry: A thorough investigation. *CEB*, *11*, 41–60. https://doi.org/10.5130/AJCEB.v11i1.1841

Trochim, W.M.K. (2006) *Research Methods: The Essential Knowledge Base*. Belmont, CA: Cengage Learning.

Tschickardt, T., Krause, D., and Akkermann, J. (2018). Development and evaluation of BIM-4-D-methods at example project Gateway Gardens lot 2, Frankfurt/Main. *Bautechnik*, *95*(7), 453–462. https://doi.org/10.1002/bate.201800026

Tserng, H. P., Yin, S. Y., Dzeng, R. J., Wou, B., Tsai, M. D., and Chen, W. Y. (2009). A study of ontology-based risk management framework of construction projects through project life cycle. *Automation in Construction*, *18*(7), 994–1008.

Tsoukalas, V. D., and Fragiadakis, N. G. (2016). Prediction of occupational risk in the shipbuilding industry using multivariable linear regression and genetic algorithm analysis. *Safety Science*, *83*, 12–22. https://doi.org/10.1016/j.ssci.2015.11.010

Tveit, M., and Gjerde, K. (2018). Using building information modelling for planning a high-speed rail project in Norway. *Civil Engineering*, *171*(3), 121–128. https://doi.org/10.1680/jcien.17.00039

Uhrin, M., Schultheis, T., Tucker, N. (2014). Stepney Green Cavern: Design Concepts and Performance of SCL Lining. *Crossrail Ltd*. Available at: https://learninglegacy.crossrail.co.uk/documents/stepney-green-cavern-design-concepts-and-performance-of-scl-lining/.

UK Government. (2021). *Approved Documents*. Department for Levelling Up, Housing and Communities. Available at: https://www.gov.uk/government/collections/approved-documents/.

Ullah, K., Lill, I., and Witt, E. (2019). An overview of BIM adoption in the construction industry: Benefits and barriers. In *10th Nordic Conference on Construction Economics and Organization* (Vol. 2) (pp. 297–303). London: Emerald Publishing Limited.

Ur-Rahman, N., and Harding, J. A. (2012). Textual data mining for industrial knowledge management and text classification: A business oriented approach. *Expert Systems with Applications*, *39*, 4729–4739. https://doi.org/10.1016/j.eswa.2011.09.124

Vijayarani, S., Ilamathi, M. J., and Nithya, M. (2015). Preprocessing techniques for text mining: An overview. *International Journal of Computer Science and Communication Networks*, *5*(1), 7–16.

Vilventhan, A., and Rajadurai, R. (2020). 4D bridge information modelling for management of bridge projects: A case study from India. *Built Environment Project and Asset Management*, *10*(3), 423–435. https://doi.org/10.1108/BEPAM-05–2019–0045

Waddoups, C. J. (2014). Union coverage and work-related training in the construction industry. *ISLR Review*, *67*(2), 532–555.

Wakefield, R. L., and Bean, D. F. (2004). Applications of text mining in large-scale databases. *International Journal of Data Mining and Knowledge Discovery*, 8(2), 147-165. https://doi.org/10.1023/B:DMKD.0000037041.22449.90

Waldman, Y. (2023) *Transportation Capital Projects Delivery Study, SFCTA*. Available at: https://www.sfcta.org/project-delivery#panel-overview (Accessed: 14 August 2023).

Wang, H., Zhang, L., and Gao, Y., (2020). Addressing cultural differences and ownership issues in the use of BIM in construction. *Automation in Construction*, 112, pp.103104.

Wang, H., Zheng, R., and Xiong, F. (2016). Application of BIM-based 5D construction technique to Chenglang Bridge. In *Proceedings of the 2016 5th International Conference on Sustainable Energy and Environment Engineering (ICSEEE, 2016)*. https://doi.org/10.2991/icseee-16.2016.6

Wang, J., and Doan, M. A. (2019). The ant empire: Fintech media and corporate convergence within and beyond Alibaba. *The Political Economy of Communication*, *6*(2), 25–37.

Wang, J., Yuan, Z., He, Z., Zhou, F., and Wu, Z. (2021). Critical factors affecting team work efficiency in BIM-Based collaborative design: An empirical study in China. *Buildings*, *11*(10), 486.

Wang, N., Satola, D., Wiberg, A. H., Liu, C., and Gustaysen, A. (2020). Reduction strategies for greenhouse gas emissions from high-speed railway station buildings in a cold climate zone of China. *Sustainability*, *2*(5), Article 1704. https://doi.org/10.3390/su12051704

Wang, X. (2012). BIM handbook: A guide to building information modelling for owners, managers, designers, engineers and contractors. *Construction Economics and Building*, *12*(3), 101–102. https://doi.org/10.5130/AJCEB.v12i3.2749C

Wang, X., Chong, W.K., Li, Q., and Lee, K.H., (2020). The Impact of Intellectual Property on BIM Implementation in Construction Projects. *Journal of Management in Engineering*, 36(1), p.04019042.

Wankhade, M., Rao, A. C. S., and Kulkarni, C. (2022). A survey on sentiment analysis methods, applications, and challenges. *Artificial Intelligence Review*, *55*(7), 5731–5780.

Weischedel, R., Palmer, M., and Kummerfeld, J. K. (2013). *OntoNotes: A large-scale corpus for linguistic annotations*. In Proceedings of the 6th Linguistic Annotation Workshop (pp. 25-32). Association for Computational Linguistics. https://doi.org/10.3115/1614527.1614533

Weiss, S. M., Indurkhya, N., Zhang, T., and Damerau, F. (2010). *Text mining: Predictive methods for analyzing unstructured information*. Cham: Springer Science and Business Media.

Wen, Q.-J., Ren, Z.-J., Lu, H., and Wu, J.-F. (2021). The progress and trend of BIM research: A bibliometrics-based visualization analysis. *Automation in Construction*, *124*, Article 103558. https://doi.org/10.1016/j.autcon.2021.103558.

Wen, X. and Xiao, L. (2020). Machine learning and NLP embedded BIM model for smart construction. *Automation in Construction*, 119, p.103318.

Wickersham, J. (2009). *Legal and business implications of building information modelling (BIM) and integrated project delivery (IPD).* Cambridge, MA: Harvard University Press.

Wiedemann, G. (2013). Opening up to Big Data: Computer-assisted analysis of textual data in social sciences. *Historical Social Research*, *38*(4), 332–357.

Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J., Appleton, G., Axton, M., Baak, A., and et al. (2016) *'The FAIR Guiding Principles for scientific data management and stewardship*', Scientific Data, 3, p. 160018. https://doi.org/10.1038/sdata.2016.18

Williams, M. (2018) The Philosophical Foundations of Mixed Methods Research. *International Journal of Social Research Methodology*, 21(6), pp. 619-632.

Williams, T. P., and Gong, J. (2014). Predicting construction cost overruns using text mining, numerical data and ensemble classifiers. *Automation in Construction*, *43*, 23–29. https://doi.org/10.1016/j.autcon.2014.02.014

Wrigley, J., Carden, V., and von Isenburg, M. (2019). Bibliometric mapping for current and potential collaboration detection. *Journal of the Medical Library Association*, *107*(4), 597–600. https://doi.org/10.5195/jmla.2019.764

Wu, H., Zuo, J., Zillante, G., Wang, J., and Yuan, H. (2019). Construction and demolition waste research: A bibliometric analysis. *Architectural Science Review*, *62*, 354–365.

Wu, Z., Wang, L., Fu, Z., Zhu, L., Dou, F., and Xu, P. (2018). VR+BIM: Perception and design optimization of highway. *International Conference on Virtual Reality and Visualization (ICVRV)* (pp. 164–165). https://doi.org/10.1109/ICVRV.2018.00059

Xiang, C., Wang, Y. B., and Liu, H. (2017). A scientometrics review on nonpoint source pollution research. *Ecological Engineering*, *99*, 400–408.

Xiong, Y., Lu, Y. and Rojas, E., (2017). Intellectual Property Protection Challenges in the BIM Environment. *Journal of Computing in Civil Engineering*, 31(4), p.04017016.

Xu, S. and Cai, H. (2021). Text mining-enabled BIM framework for predictive maintenance. *Journal of Computing in Civil Engineering*, 35(3), pp.04021009.

Xue, C. (2019). Research on the investment management of railway project construction's whole life cycle based on BIM. *Journal of Railway Engineering Society*, *36*, 103–107.

Yalcinkaya, M., and Singh, V. (2015). Patterns and trends in building information modelling (BIM) research: A latent semantic analysis. *Automation in Construction*, *59*, 68–80.

Yan, W., Zhang, J., and Huang, W. (2022). Systematic review of text mining in AEC industry. *Construction Innovation*, 22(3), pp.540-565.

Yarmohammadi, S., Pourabolghasem, R., and Castro-Lacouture, D. (2017). Mining implicit 3D modelling patterns from unstructured temporal BIM log text data.

Automation in Construction, 81, 17–24. https://doi.org/10.1016/j.autcon.2017.04.012

Yi, S., and Nie, L. (2016). Digital railway location system based on virtual geographic environment. *Journal of Southwest Jiaotong University*, *51*, 373–380. https://doi.org/10.3969/j.issn.0258–2724.2016.02.016

Yi, W., and Chan, A. P. C. (2014). Critical review of labor productivity research in construction journals. *Journal of Management in Engineering*, *30*(2), 214–225. https://doi.org/10.1061/(ASCE)ME.1943-5479.0000194

Yin, R.K. (1991) *Qualitative Research from Start to Finish*. New York, NY: The Guilford Press.

Yin, R.K. (2003) *Applications of Case Study Research*. 2nd ed. Thousand Oaks, CA: SAGE Publications.

Yin, R.K. (2006) *Mixed Methods Research: A Researcher's Guide*. Thousand Oaks, CA: SAGE Publications.

Yin, R.K. (2014) *Case Study Research: Design and Methods*. 5th ed. Thousand Oaks, CA: SAGE Publications.

Yin, R.K. (2018) *Case Study Research and Applications: Design and Methods*. 6th ed. Thousand Oaks, CA: SAGE Publications.

Young, N. W., Jones, S. A., Bernstein, H. M., and Gudgel, J. (2008). *SmartMarket report on building information modelling (BIM): Transforming design and construction to achieve greater industry productivity*. Washington, DC: McGraw-Hill Construction.

Yu, W., and Hsu, J. (2013). Content-based text mining technique for retrieval of CAD documents. *Automation in Construction*, *31*, 65–74. https://doi.org/10.1016/j.autcon.2012.11.037

Yu, W., and Skibniewski, M. J. (1999). Quantitative constructability analysis with a neuro-fuzzy knowledge-based multi-criterion decision support system. *Automation in Construction*, *8*, 553–565. https://doi.org/10.1016/S0926–5805(98)00105–8

Yusuf, J. E. W., and Jordan, M. M. (2012). Effective popular financial reports: The citizen perspective. *Journal of Government Financial Management*, *61*(4), 44–49.

Zade, J. G., and Noori, R. (2008). Prediction of municipal solid waste generation by use of artificial neural network: A case study of Mashhad. *International Journal of Environmental Research*, *2*(1), 13–22.

Zaed, O., Chen, Z., and Bareka, A. (2022). A future-proof curriculum framework for BIM excellence in architectural education. *In Innovations, Disruptions and Future Trends in the Global Construction Industry.* 25/11/2022. Routledge, UK.

Zaed, O., Chen, Z., and Bareka, A. (2022). An educational contribution to applicable digital twins through BIM integrated curriculum development in architecture.

Proceedings of the International Postgraduate Research Conference (IPGRC). 04-06/04/2022, University of Salford, Salford, UK.

Zaed, O., Chen, Z., Grant, M., Dimitrijevic, B., and Bareka, A. (2021). Developing a BIM integrated curriculum framework for undergraduate architectural education in Libya. *Proceedings of the 2nd National Conference for Developing Higher Education Institutions in Libya.* 07/10/2021, Bani Walid University, Bani Walid, Libya.

Zeng, S. X., Ma, H. Y., Lin, H., Zeng, R. C., and Tam, V. W. Y. (2015). Social responsibility of major infrastructure projects in China. *International Journal of Project Management*, *33*, 537–548. https://doi.org/10.1016/j.ijproman.2014.07.007

Zeng, W., Peng, A. C., and Zhao, J. (2023). When friendship is not mutual: The influence of network centrality incongruence on leadership emergence and organizational identification. *Journal of Leadership and Organizational Studies*, *30*(1), 11–24.

Zhai, X., Li, Z., Gao, K., Huang, Y., Lin, L., and Wang, L. (2015). Research status and trend analysis of global biomedical text mining studies in recent 10 years. *Scientometrics*, *105*, 509–523.

Zhang, F., Fleyeh, H., Wang, X., and Lu, M. (2019). Construction site accident analysis using text mining and natural language processing techniques. *Automation in Construction*, *99*, 238–248.

Zhang, K., Erfani, A., Beydoun, O. and Cui, Q., (2022). Procurement Benchmarks for Major Transportation Projects. *Transportation Research Record, 2676*(11), pp.363-376.

Zhang, L., and Dong, L. (2019). Application study on building information model (BIM) standardization of Chinese engineering breakdown structure (EBS) coding in life cycle management processes. *Advances in Civil Engineering*, *2019*, Article 1581036. https://doi.org/10.1155/2019/1581036

Zhang, M. and El-Gohary, N. (2017). Compliance checking in construction regulation using text mining. *Journal of Computing in Civil Engineering*, 31(4), pp.04017015.

Zhang, Y., Huang, C., and Zhu, C. (2019). Research on the application of railway engineering project management system based on BIM. *Journal of Railway Engineering Society*, *36*, 98–103.

Zhang, Y., Zhong, D., Wu, B., Guan, T., Yue, P., and Wu, H. (2018). 3D parametric modelling of complex geological structures for geotechnical engineering of dam foundation based on T-splines. *Computer-Aided Civil and Infrastructure Engineering*, *33*, 545–570. https://doi.org/10.1111/mice.12343

Zhao, L., Liu, Z., and Mbachu, J. (2019). Highway alignment optimization: An integrated BIM and GIS approach. *ISPRS International Journal of Geo-Information*, *8*(4), Article 172. https://doi.org/10.3390/ijgi8040172

Zhao, X., Zuo, J., Wu, G., and Huang, C. (2018). A bibliometric review of green building research 2000-2016. *Architectural Science Review*, *62*(1), 74–88. https://doi.org/10.1080/00038628.2018.1485548

Zhaohui, W., Wang, L., Fu, Z., Zhu, L., and Dou, F. (2018). VR+ BIM: Perception and design optimization of highway. In *2018 International Conference on Virtual Reality and Visualization (ICVRV)* (pp. 164–165). New York, NY: IEEE. https://doi.org/10.1109/ICVRV.2018.00059

Zhen-Zhong, H., Leng, S., Lin, J.-R., Li, S.W., and Xiao, Y.-Q. (2022). Knowledge extraction and discovery based on BIM: A critical review and future directions. *Archives of Computational Methods in Engineering*, *29*, 335–356. https://doi.org/10.1007/s11831-021-09576-9

Zhong, B., Gan, C., Luo, H., and Xing, X. (2018). Ontology-based framework for building environmental monitoring and compliance checking under BIM environment. *Building and Environment*, *141*, 127–142. https://doi.org/10.1016/j.buildenv.2018.05.046

Zhong, R., Xu, X., and Zhang, J. (2019). Multilingual text analysis for BIM in construction. *Automation in Construction*, 103, pp.52-62.

Zhou, H., Wang, H., and Zeng, W. (2018). Smart construction site in mega construction projects: A case study on island tunneling project of Hong Kong-Zhuhai-Macao Bridge. *Frontiers of Engineering Management*, *5*(1), 78–87. https://doi.org/10.15302/J-FEM-2018075

Zhou, J. (2020). Research on the development process and deliverables of IFC Railway Standard. *Journal of Railway Engineering Society*, *37*, 92–98.

Zhou, Q., Liao, C., and Wu, Y., (2013). A study of the role of data management in construction projects: Unstructured data analysis for improved decision-making. *Construction Management and Economics*, 31(9), pp.911-920.

Zhou, S., Ng, S., Lee, S., Xu, F., and Yang, Y. (2020). A domain knowledge incorporated text mining approach for capturing user needs on BIM applications. *Engineering, Construction and Architectural Management*, *27*(2), 458–482. https://doi.org/10.1108/ECAM-02-2019-0097

Zhou, W., Zhang, L. and Xie, X., (2013). Challenges and Opportunities for the Application of BIM in the Construction Industry. *Automation in Construction*, 35, pp. 8-17.

Zhou, Y., Zhang, L., and Zhang, H. (2019). Text mining and big data analytics in construction project management. *Automation in Construction*, 108, 124-132.

Zhou, Y., Zhao, Y., and Tan, J. (2019). Needs analysis of BIM in construction based on TM. Journal of Civil Engineering and Management, 25(3), pp.220-231.

Zidane, Y. J.-T., Johansen, A., and Ekambaram, A. (2013). Megaprojects: Challenges and lessons learned. *Procedia - Social and Behavioral Sciences*, *74*, 349–357. https://doi.org/10.1016/j.sbspro.2013.03.041

Zohuri, B., Moghaddam, M., and Mossavar-Rahmani, F. (2022). Business resilience system integrated Artificial Intelligence system. *International Journal of Theoretical and Computational Physics*, *3*(1), 1–7.

Zou, Y., Kiviniemi, A., and Jones, S. W. (2016). Developing a tailored RBS linking to BIM for risk management of bridge projects. *Engineering, Construction and Architectural Management*, *23*, 727–750. https://doi.org/10.1108/ECAM-01-2016-0009.

Appendices

Appendix 1: Dataset of Chosen articles

The dataset has already been published on Mendeley and is now live for access. It has undergone a review process to maintain high levels of quality on Mendeley Data and to ensure that it meets the necessary standards. The system has also performed checks to prevent spam. The table below displays information from various fields of the retrieved data obtained from dedicated databases. It includes the DOI for each article, the author(s), and their respective titles. Additionally, the dataset generated from web scraping activities is accessible via the following DOI:
https://doi.org/10.17632/5hcg7tddc4.3

No	Journals	Year	Titles of Articles
1	Energy and Buildings	2014	Monitoring thermal comfort in subways using building information modeling
2	Proceedings of the Institution of Civil Engineers: Civil Engineering	2014	Four-dimensional modelling on Tottenham court road station, London, UK
3	Proceedings of the Institution of Civil Engineers: Civil Engineering	2014	Building information modelling in practice: transforming Gatwick airport, UK
4	Automation in Construction	2014	BIM-based framework for managing performance of subway stations
5	Journal of Construction Engineering and Management	2014	Highway alignment construction comparison using object-oriented 3D visualization modeling
6	Buildings	2015	Assessing embodied energy and greenhouse gas emissions in infrastructure projects
7	Practice Periodical on Structural Design and Construction	2015	Implementing BIM on infrastructure: comparison of two bridge construction projects

8	Visualization in Engineering	2015	Application of BIM coordination technology to HSR Changhua station
9	Sensors (Switzerland)	2015	Development of a 3D underground cadastral system with indoor mapping for as-built BIM: the case study of Gangnam subway station in Korea
10	Engineering, Construction and Architectural Management	2016	Developing a tailored RBS linking to BIM for risk management of bridge projects
11	Malaysian Construction Research Journal	2016	An exploratory study: Building Information Modelling Execution Plan (BEP) procedure in mega construction projects
12	Journal of Performance of Constructed Facilities	2016	Underground pipeline management based on road information modeling to assist in road management
13	Journal of Bridge Engineering	2016	Bridge Information Modeling for inspection and evaluation
14	Journal of Civil Engineering and Management	2016	BIM-based risk identification system in tunnel construction
15	Automation in Construction	2016	Managing construction schedule by telepresence: integration of site video feed with an active nd CAD simulation
16	Journal of Civil Structural Health Monitoring	2016	Defining a conceptual framework for the integration of modelling and advanced imaging for improving the reliability and efficiency of bridge assessment
17	KSCE Journal of Civil Engineering	2016	Development of an IFC-based data schema for the design information representation of the NATM tunnel
18	Advances in Engineering Software	2016	Construction and facility management of large MEP projects using a multi-scale building information model
19	Visualization in Engineering	2016	Optimization of a hybrid tower for onshore wind turbines by building information modeling and prefabrication techniques
20	Proceedings of the Institution of Civil Engineers: Municipal Engineer	2016	Highway infrastructure and building information modelling in UK

21	Cluster Computing	2016	Clustering and ontology-based information integration framework for surface subsidence risk mitigation in underground tunnels
22	Facilities	2017	Leveraging BIM and big data to deliver well maintained highways
23	Open Construction and Building Technology Journal	2017	The information requirements for transportation industrys facilities management based on BIM
24	Agro Food Industry Hi-Tech	2017	The application of BIM technology in simulation modeling of concrete dam construction
25	Transportation Research Part A: Policy and Practice	2017	Off the rails: the cost performance of infrastructure rail projects
26	Transportation Research Part A: Policy and Practice	2017	Light rail transit cost performance: opportunities for future-proofing
27	Journal of Computing in Civil Engineering	2017	Management of collaborative BIM data by federating distributed BIM models
28	Proceedings of the Institution of Civil Engineers: Civil Engineering	2017	Crossrail project: building a virtual version of London’s Elizabeth line
29	Smart Structures and Systems	2017	Development of BIM-based bridge maintenance system for cable-stayed bridges
30	Journal of Construction Engineering and Management	2017	BIM-based collaboration platform for the management of EPC projects in hydropower engineering
31	International Journal of Sustainable Development and Planning	2018	Using BIM models for the design of large rail infrastructure projects: key factors for a successful implementation
32	Structural Engineering International	2018	Three-dimensional information delivery for design and construction of prefabricated bridge piers
33	Proceedings of the Institution of Civil Engineers: Civil Engineering	2018	Using building information modelling for planning a high-speed rail project in Norway

34	Journal of Computing in Civil Engineering	2018	Automatic as-built BIM creation of precast concrete bridge deck panels using laser scan data
35	Automation in Construction	2018	An automated safety risk recognition mechanism for underground construction at the pre-construction stage based on BIM
36	Tunnelling and Underground Space Technology	2018	An integrated system framework of building information modelling and geographical information system for utility tunnel maintenance management
37	Engineering, Construction and Architectural Management	2018	Mobile BIM implementation and lean interaction on construction site: a case study of a complex airport project
38	Engineering, Construction and Architectural Management	2018	Exploring the BIM and lean synergies in the Istanbul grand airport construction project
39	Sustainability (Switzerland)	2018	Benefit-cost analysis of Building Information Modeling (BIM) in a railway site
40	Journal of Computing in Civil Engineering	2018	Structural performance monitoring using a dynamic data-driven BIM environment
41	Journal of Performance of Constructed Facilities	2018	Video to brim: automated 3D as-built documentation of bridges
42	Architecture and Engineering	2019	7d BIM for sustainability assessment in design processes: a case study of design of alternatives in severe climate and heavy use conditions
43	Journal of Civil Engineering and Management	2019	Research on management and application of tunnel engineering based on BIM technology
44	Infrastructures	2019	A case study of BIM implementation in rail track rehabilitation
45	Journal Europeen des Systemes Automatises	2019	A safe evacuation mode for ultradeep underground space in urban rail transit stations
46	Advances in Civil Engineering	2019	Application study on Building Information Model (BIM) standardization of Chinese Engineering

			Breakdown Structure (EBS) coding in life cycle management
47	Buildings	2019	Challenges and enablers in BIM-enabled digital transformation in mega projects: the Istanbul new airport project case study
48	Proceedings of the Institution of Civil Engineers: Civil Engineering	2019	Chaoyang bridge, China: a double-deck cable-stayed bridge with corrugated steel webs
49	KSCE Journal of Civil Engineering	2019	Risk information management for bridges by integrating risk breakdown structure into 3d/4d BIM
50	Proceedings of the Institution of Civil Engineers: Civil Engineering	2019	Using building information modelling for a commercial building in Beijing, China
51	Journal of Computing in Civil Engineering	2019	Computationally efficient simulation in urban mechanized tunnelling based on multilevel BIM models
52	Smart and Sustainable Built Environment	2019	Overview: the opportunity of BIM in railway
53	Proceedings of the Institution of Civil Engineers: Municipal Engineer	2019	Applications of building information modelling for planning and delivery of rapid transit
54	Computers in Industry	2019	Dynamic BIM-augmented UAV safety inspection for water diversion project
55	Journal of Cleaner Production	2019	Digital twin aided sustainability-based lifecycle management for railway turnout systems
56	Sustainability (Switzerland)	2019	Development of a bridge management system based on the building information modeling technology
57	Structure and Infrastructure Engineering	2019	Improved visualization of infrastructure monitoring data using building information modeling
58	Water (Switzerland)	2019	Three-dimensional numerical simulation of dam discharge and flood routing in wudu reservoir
59	Automation in Construction	2019	Automatic creation of as-is building information model from single-track railway tunnel point clouds

60	Tunnelling and Underground Space Technology	2019	Predictions of settlement risk induced by tunnelling using BIM and 3d visualization tools
61	Automation in Construction	2019	An integrated underground utility management and decision support based on BIM and GIS
62	Journal of Computing in Civil Engineering	2019	BIM+ topology diagram-driven multiutility tunnel emergency response method
63	Advances in Civil Engineering	2020	BIM adoption for facility management in urban rail transit: an innovation diffusion theory perspective
64	Advances in Civil Engineering	2020	Prediction of highway tunnel pavement performance based on digital twin and multiple time series stacking
65	Journal of Advanced Transportation	2020	Collaboration-based BIM model development management system for general contractors in infrastructure projects
66	Environmental Engineering and Management Journal	2020	Building information modeling application for groundwater recharge: development of multiple structures
67	CMES - Computer Modeling in Engineering and Sciences	2020	Building information modeling-based secondary development system for 3D modeling of underground pipelines
68	Automation in Construction	2020	Model-based quality assurance in railway infrastructure planning
69	Journal of Management in Engineering	2020	Flexible 3d model partitioning system for nd-based BIM implementation of alignment-based civil infrastructure
70	Sustainability (Switzerland)	2020	Reduction strategies for greenhouse gas emissions from high-speed railway station buildings in a cold climate zone of China
71	Sustainability (Switzerland)	2020	Sustainability-based lifecycle management for bridge infrastructure using 6D BIM
72	Tunnelling and Underground Space Technology	2020	Development and application of a specification-compliant highway tunnel facility management system based on BIM

73	Tunnelling and Underground Space Technology	2020	A BIM-based framework for operation and maintenance of utility tunnels
74	Automation in Construction	2020	Parametric modeling and structure verification of asphalt pavement based on BIM-ABAQUS
75	Engineering, Construction and Architectural Management	2020	Brim and UAS for bridge inspections and management
76	Automation in Construction	2020	Bridge damage: detection, IFC-based semantic enrichment and visualization
77	Automation in Construction	2020	Semi-automated generation of parametric BIM for steel structures based on terrestrial laser scanning data
78	Frontiers in Built Environment	2020	A benefits prioritization analysis on adopting BIM systems against major challenges in megaproject delivery
79	Infrastructures	2020	BIM approach for modeling airports terminal expansion
80	Applied Sciences (Switzerland)	2020	Design of railway track model with three-dimensional alignment based on extended industry foundation classes
81	Computer Communications	2020	Automatic underground space security monitoring based on BIM
82	Automation in Construction	2020	Integrating three-dimensional road design and pavement structure analysis based on BIM
83	Frontiers in Built Environment	2020	System identification-enhanced visualization tool for infrastructure monitoring and maintenance
84	Transportation Research Record	2020	Building information modeling implementation framework for smart airport life cycle management
85	International Journal of Sustainable Development and Planning	2020	Visualization of foundation evaluation for urban rail transit based on CGB technology integration
86	Sustainability (Switzerland)	2020	Evaluation of space service quality for facilitating efficient operations in a mass rapid transit station

87	Electronic Journal of Information Systems in Developing Countries	2020	Implementation of building information modeling in Vietnamese infrastructure construction: a case study of institutional influences on a bridge project
88	Built Environment Project and Asset Management	2020	4D Bridge Information Modelling for management of bridge projects: a case study from India
89	Journal of Engineering, Design and Technology	2020	Research on dynamic data monitoring of steel structure building information using BIM
90	Advanced Engineering Informatics	2020	State of the art in damage information modeling for RC bridges a literature review
91	Sustainability (Switzerland)	2020	Digital twin aided sustainability and vulnerability audit for subway stations
92	Sensors (Switzerland)	2020	A framework for an indoor safety management system based on digital twin
93	Applied Sciences (Switzerland)	2020	A master digital model for suspension bridges
94	Sustainability (Switzerland)	2020	Assessing the economic risk of building damage due to the tunnelling-induced settlement using Monte Carlo simulations and BIM
95	Automation in Construction	2020	BIM-based environmental impact assessment for infrastructure design projects
96	International Journal of Systems Assurance Engineering and Management	2020	BIM based risk management for Design Bid Build (DBB) design process in the United Arab Emirates: a conceptual framework
97	Underground Space (China)	2020	BIM for the underground an enabler of trenchless construction
98	Proceedings of the Institution of Civil Engineers-Municipal Engineer	2021	Implementation of building information modeling standards for the shanghai metro, China
99	International Journal of Systems Assurance	2021	Virtual construction technology of tunnel engineering based on BIM platform and measuring robot

	Engineering and Management		
100	KSCE Journal of Civil Engineering	2021	A case study on BIM object-based earned value and process management in highway construction
101	KSCE Journal of Civil Engineering	2021	A BIM based approach for structural health monitoring of bridges
102	International Journal of Pavement Engineering	2021	Pavement condition information modelling in an I-BIM environment
103	Construction Innovation	2021	BIM-based mixed-reality application for bridge inspection and maintenance
104	Advances in Civil Engineering	2021	Safety management system prototype/framework of deep foundation pit based on BIM and IoT
105	Advances in Civil Engineering	2021	Crane lifting optimization and construction monitoring in steel bridge construction project based on BIM and UAV
106	Georisk	2021	3d spatiotemporal risk assessment analysis of the tunnelling-induced settlement in an urban area using analytical hierarchy process and BIM
107	Advances in Civil Engineering	2021	Design and application of risk early warning system for subway station construction based on building information modeling real-time model
108	Journal of Advanced Transportation	2021	Procedural modeling-based BIM approach for railway design
109	Structural Engineering International	2021	Study of bridge scheme creation based on BIM technology
110	Neural Computing and Applications	2021	Early warning and real-time control of construction safety risk of underground engineering based on building information modeling and Internet of Things
111	Advances in Civil Engineering	2021	Automated reconstruction of parametric BIM for bridge based on terrestrial laser scanning data
112	Advances in Civil Engineering	2021	Intelligent early warning system for construction safety of excavations adjacent to existing metro tunnels

113	Journal of Asian Architecture and Building Engineering	2021	Visualization and monitoring information management of bridge structure health and safety early warning based on BIM
114	Tunnelling and Underground Space Technology	2021	From digital models to numerical analysis for mechanised tunnelling: a fully automated design-through-analysis workflow
115	Journal of Air Transport Management	2021	Opportunities in airport pavement management: integration of BIM, the IoT and DLT
116	Arabian Journal of Geosciences	2021	Research on dynamic data monitoring of marine bridge steel structure building information based on BIM model
117	Tunnelling and Underground Space Technology	2021	BIM, machine learning and computer vision techniques in underground construction: current status and future perspectives
118	Tunnelling and Underground Space Technology	2021	I-BIM based approach for geotechnical and numerical modelling of a conventional tunnel excavation
119	Tunnelling and Underground Space Technology	2021	BIM implementation in handover management for underground rail transit project: a case study approach
120	Sustainability (Switzerland)	2021	Digital twin aided vulnerability assessment and risk-based maintenance planning of bridge infrastructures exposed to extreme conditions
121	Engineering, Construction and Architectural Management	2021	Airport project delivery within BIM-centric construction technology ecosystems
122	Automation in Construction	2021	BIM-based safety design for emergency evacuation of metro stations
123	Computers, Environment and Urban Systems	2021	Topological integration of BIM and geospatial water utility networks across the building envelope
124	Journal of Construction Engineering and Management	2021	Implementation of a BIM-FM platform at an international airport project: case study

125	Sustainability (Switzerland)	2021	Development of BIM-based bridge maintenance system considering maintenance data schema and information system
126	Automation in Construction	2021	A parametric BIM approach to foster bridge project design and analysis
127	Applied Sciences (Switzerland)	2021	BIM-GIS-based integrated framework for underground utility management system for earthwork operations
128	Applied Sciences (Switzerland)	2021	BIM and GIS applications in bridge projects: a critical review
129	Buildings	2021	Utilization of building information modeling for arranging the structural kingposts
130	Automation in Construction	2021	Building Information Modelling (BIM) application for an existing road infrastructure
131	Applied Sciences (Switzerland)	2021	Case study of solar photovoltaic power-plant site selection for infrastructure planning using a BIM-GIS-based approach
132	Automation in Construction	2021	An IFC schema extension for BIM-based description of wastewater treatment plants
133	Journal of Transportation Engineering Part B: Pavements	2021	3D visualization of airport pavement quality based on BIM and WEBGL integration
134	Energies	2021	A digital information model framework for UAS-enabled bridge inspection
135	Journal of Civil Engineering and Management	2021	A BIM-based identification and classification method of environmental risks in the design of Beijing subway
136	PLoS ONE	2021	Integrated BIM and VR to implement IPD mode in transportation infrastructure projects: system design and case application
137	Automation in Construction	2021	BIM-based traffic analysis and simulation at road intersection design
138	Journal of Civil Structural Health Monitoring	2021	Building information modeling-based bridge health monitoring for anomaly detection under complex loading conditions using artificial neural networks
139	KSCE Journal of Civil Engineering	2021	Implementation of BIM + WEBGIS based on extended IFC and batched 3d tiles data: an

			application in RCC gravity dam for republication of design change
140	Journal of Engineering, Design and Technology	2021	Analysis framework for the interactions between Building Information Modelling (BIM) and lean construction on construction mega-projects
141	Transportation Research Record	2021	Building information modeling for bridges and structures: outcomes and lessons learned from the steel bridge industry
142	Journal of Building Engineering	2021	Cim-based modeling and simulating technology roadmap for maintaining and managing Chinese rural traditional residential dwellings
143	Journal of Construction Engineering and Management	2022	Computational geometric approach for BIM semantic enrichment to support automated underground garage compliance checking
144	Tunnelling and Underground Space Technology	2022	Multi-Lod BIM for underground metro station: interoperability and design-to-design enhancement
145	Journal of Computing in Civil Engineering	2022	Framework for developing IFC-based 3D documentation from 2D bridge drawings
146	Journal of Pipeline Systems Engineering and Practice	2022	Integrated augmented reality and cloud computing approach for infrastructure utilities maintenance

Appendix 2: Relevance of the Key Technical Terms to the BIMTAPE Framework

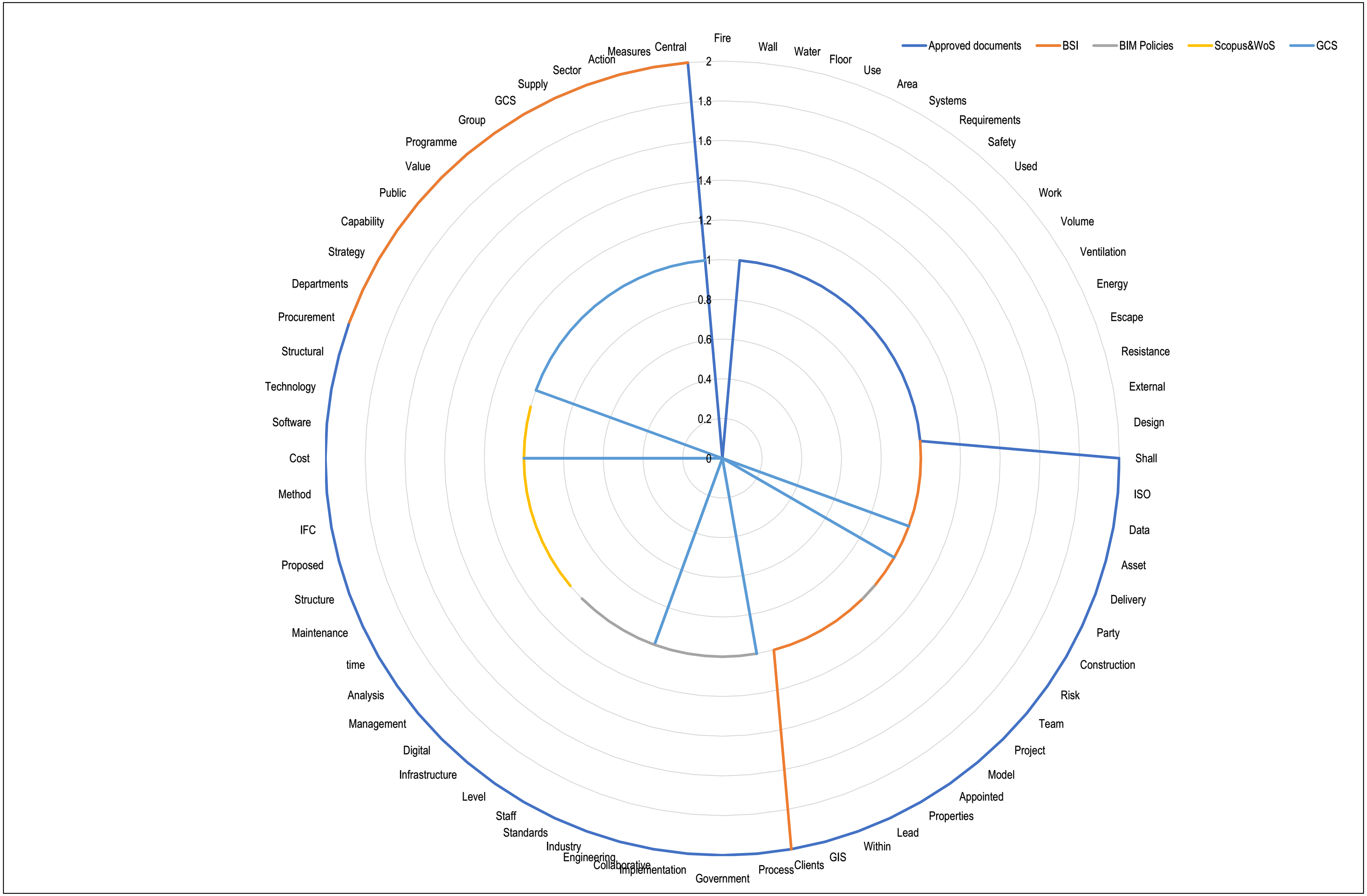
Table 7.1 describes the relevance of key technical terms to the BIMTAPE framework.

Table 7.1: The relevance of key technical terms to the BIMTAPE framework

Technical Terms	**Description and Relevance to BIMTAPE Framework**
AI	AI, through techniques like TM, ML, NLP, and Data Visualisation, is highly relevant to the research undertaken and the BIMTAPE framework. Using tools such as R Studio and Python, AI enables the processing of unstructured textual data, thereby enhancing decision-making, project scheduling, and resource management. Techniques such as Topic Modelling and Trend Analysis extracted from BIM-related data streamline workflows, optimise risk management, and contribute to more efficient project delivery. This integration of AI into the BIMTAPE framework supports a data-driven approach, facilitating improved project performance and collaboration, and directly aligns with the objectives of the research.
Knowledge Graph~~s~~ (KG~~s~~)	KGs play a crucial role in the BIMTAPE framework by offering a semantic structure that improves the handling of complex relationships in BIM data. KGs enhance the organisation and interlinking of diverse data sources, supporting efficient retrieval and analysis. In this research, KGs facilitate advanced decision-making, streamline data accessibility, and provide a more integrated understanding of project components. By enabling graph-based reasoning and contextual searches, KGs help uncover implicit relationships between BIM concepts, standards, and regulations, thus driving informed and efficient project delivery. Their integration is vital for addressing interdisciplinary connections and optimising construction practices, making KGs a key component of the BIMTAPE framework.
BIM Knowledge Graph (BIMKG)	The BIMKG enhances the BIMTAPE framework by linking the structured categorisation of the Periodic Table of BIM (Mordue, 2016), with dynamic insights into BIM element interactions. This integration provides a clear visualisation of how various BIM components influence one another, supporting decision-making and knowledge management in BIM pervasive project delivery. By revealing relationships across BIM elements, the BIMKG adds depth to the BIMTAPE framework, aiding in the strategic understanding and optimisation of BIM in large-scale project delivery.

Appendix 3: Node Distances in Network

Annex 3 presents an enlarged version of Figure 4.88, offering a clearer visualization of node distance distribution derived from SNA.


Approved documents
BSI
BIM Policies
Scopus&WoS
GCS
2
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
Fire
Wall
Water
Floor
Use
Area
Systems
Requirements
Safety
Used
Work
Volume
Ventilation
Energy
Escape
Resistance
External
Design
Shall
ISO
Data
Asset
Delivery
Party
Construction
Risk
Team
Project
Model
Appointed
Properties
Lead
Within
GIS
Clients
Process
Government
Implementation
Collaborative
Engineering
Industry
Standards
Staff
Level
Infrastructure
Digital
Management
Analysis
time
Maintenance
Structure
Proposed
IFC
Method
Cost
Software
Technology
Structural
Procurement
Departments
Strategy
Capability
Public
Value
Programme
Group
GCS
Supply
Sector
Action
Measures
Central