www.itcon.org - Journal of Information Technology in Construction - ISSN 1874-4753

FROM NLP TO TAXONOMY: IDENTIFYING AND CLASSIFYING KEY FUNCTIONALITY CONCEPTS OF MULTI-LEVEL PROJECT PLANNING AND CONTROL SYSTEMS

SUBMITTED: May 2024
REVISED: July 2024
PUBLISHED: December 2024
GUEST EDITORS: Vito Getuli, Farzad Rahimian, Nashwan Dawood, Pietro Capone, Alessandro Bruttini
DOI: 10.36680/j.itcon.2024.053

Moslem Sheikhkhoshkar, Research Assistant
CNRS, CRAN, Université de Lorraine, Epinal, France
moslem.sheikhkhoshkar@univ-lorraine.fr

Hind Bril El Haouzi, Professor
CNRS, CRAN, Université de Lorraine, Nancy, France
hind.el-haouzi@univ-lorraine.fr

Alexis Aubry, Associate Professor
CNRS, CRAN, Université de Lorraine, Nancy, France
alexis.aubry@univ-lorraine.fr

Farook Hamzeh, Professor
University of Alberta, Edmonton, Canada
hamzeh@ualberta.ca

Farzad Rahimian, Professor
Teesside University, Middlesbrough, UK
F.Rahimian@tees.ac.uk

SUMMARY: *Analysis of literature and industry practices in applied planning and control systems reveals a notable lack of effective processes and stakeholders' understanding regarding the optimal use of these systems. These gaps underscore the urgent need for a refined understanding and discovery of the underlying concepts of existing systems to address the complex dynamics of the planning and control domain better. Therefore, this study employed a multi-step approach using advanced text-mining techniques and expert validation to address these issues. Sentence-Bidirectional Encoder Representations from Transformers (SBERT) for semantic analysis, hierarchical clustering, and word cloud visualization were applied to classify and validate project planning and control system functionality concepts into coherent clusters. Furthermore, a robust taxonomy of functionality concepts was developed by meticulously analysing the findings as well as considering the domain experts' insights. As a result, 148 project planning and control systems' functionalities were classified into 20 coherent clusters with an average 87% alignment rate. A robust taxonomy of these functionalities was then formulated, emphasizing their importance across various scheduling levels. This taxonomy captures the complexities of project planning and control systems, facilitating informed decision-making and the integration of diverse planning and control systems to handle project complexities. The research significantly contributes to the field by clarifying the core concepts of project planning and control systems, making them more understandable and actionable for project stakeholders.*

KEYWORDS: *Natural Language Processing (NLP), Taxonomy, Functionality concepts, Planning and control systems, SBERT.*

REFERENCE: *Moslem Sheikhkhoshkar, Hind Bril El Haouzi, Alexis Aubry, Farook Hamzeh, Farzad Rahimian (2024). From NLP to Taxonomy: Identifying and Classifying Key Functionality Concepts of Multi-level Project Planning and Control Systems. Journal of Information Technology in Construction (ITcon), Special issue: 'Managing the digital transformation of construction industry (CONVR 2023)', Vol. 29, pg. 1200-1218, DOI: 10.36680/j.itcon.2024.053*

1. INTRODUCTION

Effective production planning and control are crucial for achieving project objectives in construction projects that are characterized by significant uncertainties (Nwadigo et al. 2021; Viana et al. 2017). The primary objectives of planning and control systems are to facilitate managerial decision-making, orchestrate communication and coordination among various project stakeholders, and establish benchmarks for performance measurement and evaluation (Laufer and Tucker 1987). Despite its significance, the implementation and outcomes of production planning and control often evoke dissatisfaction within the construction industry. Indeed, deficiencies in effective planning and control processes have been reported as one of the major contributors to project delays and cost overruns (Sheikhkhoshkar et al. 2023a; Viana et al. 2017). Furthermore, studies have highlighted a pervasive lack of understanding and experience among project stakeholders concerning planning and control systems (AlNasseri and Aulin 2015; Salling et al. 2023; Sheikhkhoshkar et al. 2024b).

As construction projects grow in complexity and markets evolve with increased dynamism and fragmentation, the prevalence of outsourcing and subcontracting escalates, presenting additional challenges in coordinating supply chains. These challenges necessitate managing numerous planning alternatives and divergent stakeholder interests, alongside a pervasive lack of comprehensive project understanding among participants (Viana et al. 2017; Wiendahl et al. 2005). In this regard, various single and integrated planning and control systems have been proposed in the last decade to address these challenges. Systems such as the Last Planner System (LPS) (Ballard and Tommelein 2021), Takt Time Planning (TTP) (Frandson 2019), and Location-Based Management System (LBMS) (Ghanem et al. 2022) have contributed to the body of production planning and control knowledge through enhancing communication and collaboration, standardizing workflows, improving resource utilization, optimizing scheduling and logistics, and facilitating continuous improvement (Ezzeddine et al. 2022; O. AlSehaimi et al. 2014). Furthermore, these systems have been combined with one another to augment their effectiveness and address the drawbacks inherent in each, resulting in hybrid systems such as the Building Information Modelling (BIM)-LPS-Kanban (Arayici et al. 2023; Sacks et al. 2010), Critical Chain Project Management (CCPM)-LPS-Linear Scheduling Method (LSM) (Salama et al. 2021), LBMS-LPS-Critical Path Method (CPM) (Olivieri et al. 2016), and 4DBIM-LPS-LBMS (Silveira and Costa 2023), LPS-LBMS (Seppänen et al. 2010), BIM-Kanban (Zeng et al. 2023), and others. The objectives behind the development of these approaches include improving the reliability of planning, decreasing workflow variability, increasing teamwork and communication among all stakeholders, and enhancing the understanding of the project's progress, among many others.

Additionally, advancements in control systems aspects have been pursued with the introduction of systems such as Earned Value Management (EVM) (Naderi et al. 2024; Sheikhkhoshkar et al. 2024b), lean-based control systems (Hamzeh et al. 2019), and buffer control systems (Hu et al. 2016; Zohrehvandi et al. 2021). These systems are designed to measure and monitor various aspects of projects including uncertainties, quality of capacity planning, commitments' quality, workflows' reliability and stability, project progress and performance, and the efficiency of resource allocation.

Despite the awareness of these advancements among academics and some policymakers in the construction sector, a significant gap exists in the knowledge and utilization of these systems among on-the-ground project teams, including project managers, planners, and superintendents (Salling et al. 2023; Sheikhkhoshkar et al. 2023a). This disconnect underscores the necessity of clarifying the main objectives and underlying principles of these developments to enhance the level of understanding and awareness among project teams. Such clarity will enable more effective implementation of these systems in construction projects. Moreover, there is a lack of a structured framework to classify and organize the various functionalities of the planning and control systems, making it challenging for stakeholders to determine the appropriate level of effort needed at each schedule level to effectively address specific functional requirements for a planning and control system. These gaps necessitate extraction and analysis of the underlying concepts and objectives of planning and control systems as well as developing a taxonomy to systematically classify and organize project planning and control systems' functionalities, enhancing stakeholders' understanding and utilization.

In this regard, the studies conducted by Sheikhkhoshkar et al. (2023b), Sheikhkhoshkar et al. (2023a), and Sheikhkhoshkar et al. (2024b) defined the underlying concepts and objectives of planning and control systems as their functionalities. They then extracted and analysed a holistic list of functionalities through systematic literature reviews and meta-analyses. However, the resulting list of 148 functionalities is overly extensive, making it challenging to clearly grasp the core ideas and principles of planning and control systems. Therefore, this study aims first to analyze a comprehensive list of functionalities and identify the main concepts behind them using a Natural Language Processing (NLP) approach. Next, it seeks to develop a taxonomy to simplify complex concepts and support informed decision-making. In pursuit of this aim, the outlined objectives of this research include:

1- Analyse the project planning and control system's functionalities and identify the main functionality concepts
2- Develop a taxonomy to classify the identified functionality concepts

This research contributes to the body of knowledge in production planning and control by shedding light on the underlying concepts of project planning and control systems and making them more tangible and clear for project stakeholders. Additionally, through the development of a taxonomy of functionality concepts, the research will highlight the importance of each concept at different scheduling levels. This will enable project teams to better understand their functional requirements associated with planning and control systems at each schedule level. Moreover, this study helps translate the tacit knowledge of domain experts into formalized knowledge that can be easily shared and used by practitioners. Additionally, it lays the groundwork for creating a knowledge repository that can incorporate semantic technologies to infer new knowledge and develop expert systems.

The subsequent sections of this paper are organized as follows:

An exploration of NLP within the Architecture, Engineering, and Construction (AEC) domain; the research methodology employed; the findings, analysis, and validation of the results, research discussion and implications; and finally, the conclusion, limitations and suggestions for future endeavours.

2. NLP IN AEC DOMAIN

NLP presents a viable computational methodology to process and analyze large amounts of natural language data, enabling them to understand and interpret in a valuable way (Kang et al. 2020). NLP is making significant progress in the AEC domain, addressing complex challenges such as efficient information extraction, project planning, and compliance management (Li et al. 2024). Leveraging NLP technology improves the processing of the extensive unstructured textual data embedded in construction documents, schedules, and contracts, which is crucial for improving project management and operational efficiency. One significant application of NLP in the AEC is extracting semantics from regulatory texts, where NLP can efficiently parse and interpret complex documents to aid compliance checks, risk and contract management. In this context, Moon et al. (2022) applied the Bidirectional Encoder Representations from Transformers (BERT) method in natural language processing to detect contractual risk clauses from construction specifications. Ko et al. (2021) proposed an NLP–driven model to extract and analyze contract change reasons, enhancing the precision of change order management systems. A parser was utilized by Al Qady and Kandil (2010) to extract semantic knowledge from construction contract documents to improve electronic document management capabilities.

Another critical application of NLP in the AEC sector is in project planning and scheduling. NLP models may discover and extract actionable insights from textual data in project documents, such as schedule information and task dependencies, which are fed into project management processes. This capability not only streamlines project scheduling processes but also ensures that they are more aligned with operational realities, which enhances overall project efficiency. Taking this into account, Jung and Golparvar-Fard (2023) employed NLP for both master and lookahead schedules to automatically learn and map their respective activities and tasks. This approach aimed to address the gap in alignment between schedule levels, thereby streamlining the manual reconciliation process, which is susceptible to errors and inefficiencies. Jung et al. (2024) introduced a novel NLP-based model to automate the linkage between textual descriptions of scheduled activities and ASTM Uniformat categories. This automation seeks to ease manual processes and organizational challenges in implementing 4D BIM. Key features include the automated creation of 4D BIM, mapping schedule activities to payment applications, and computer vision progress monitoring without reliance on BIM. This approach eliminates the need for manual efforts in synchronizing activity IDs with corresponding BIM elements across Virtual Design and Construction (VDC) and planning departments.

Furthermore, Zhao et al. (2020) developed a system that employs automatic project schedule checking (APSC) and NLP to extract construction methods from project schedules, thereby aiding in the automated assessment of schedule quality in the construction sector. This solution combines NLP techniques with OWL-based ontology models to derive semantic and syntactic features for construction activities. Ko et al. (2023) explored a method that employs NLP to systematically measure the similarity between project scope statements to recommend similar projects for reliable project development and planning in the preconstruction phase.

To sum up, NLP is used in the construction industry for various purposes involving textual data, ranging from compliance and risk management to project planning and scheduling. This paper also aims to apply a new application of NLP, leveraging its capabilities to discover and formalize the key underlying concepts of project planning and control systems.

3. RESEARCH METHODOLOGY

A multi-stage methodology was implemented to fulfil the objectives of this paper, as depicted in Figure 1. The methodology comprises two main processes: data collection and data analysis. In the data collection phase, a comprehensive systematic literature review on planning and control systems in construction was conducted to extract the functionalities of planning methods and control metrics. During the data analysis phase, following the data preprocessing, the Sentence- Bidirectional Encoder Representations from Transformers (SBERT) model, a specialized variant of BERT (Reimers and Gurevych 2019), was employed for vectorization and text embedding. The hierarchical clustering technique was then applied to cluster the extracted functionalities. Finally, word cloud generation, keyword extraction, and semi-structured interviews with domain experts were conducted to identify the primary concepts and construct the taxonomy of project planning and control systems' functionalities. Each facet of the research methodology is described in detail in the following subsections.

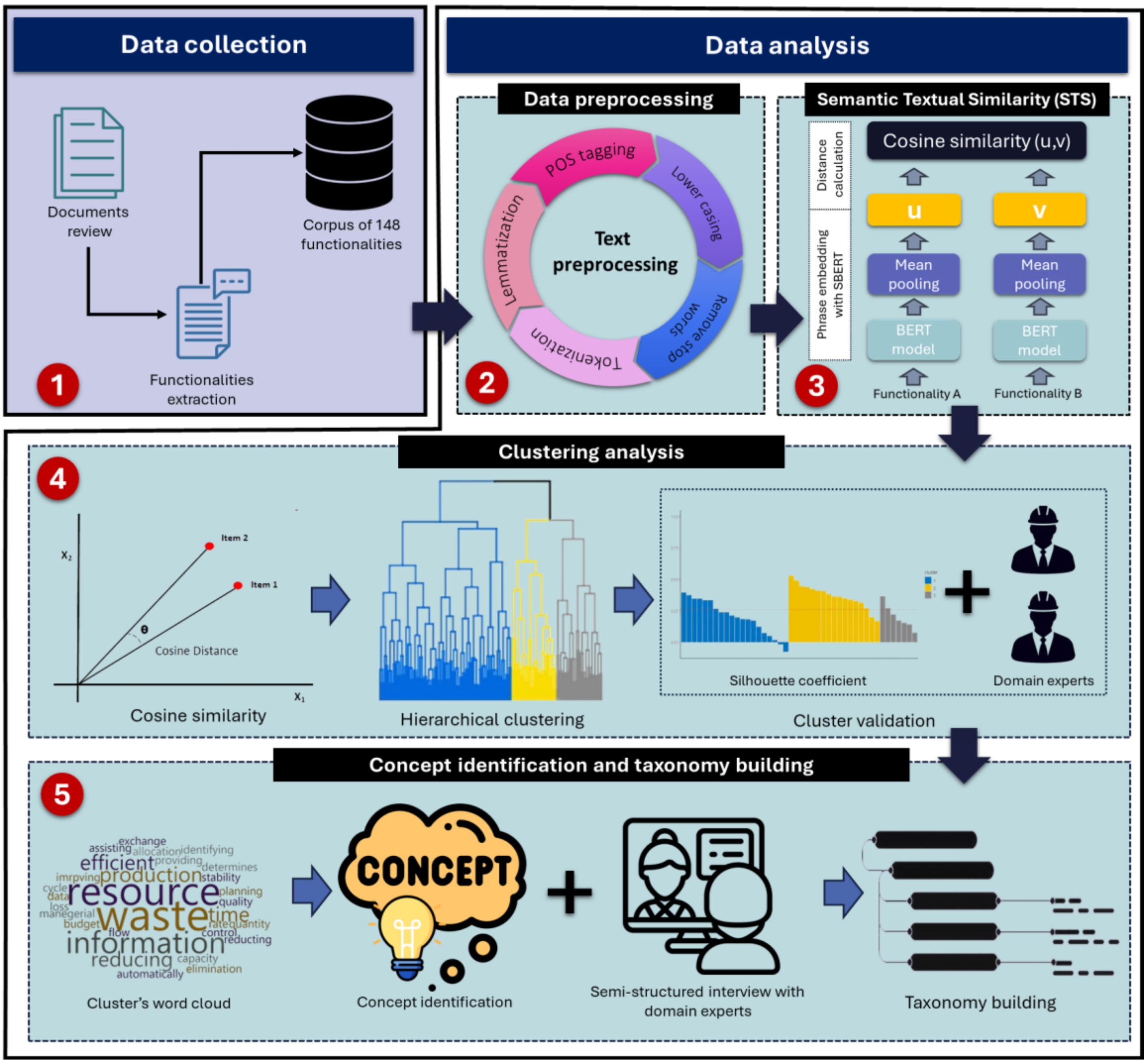


Figure 1: An overview of the adopted methodology.

3.1 Data Collection

Three systematic literature reviews and meta-analyses were carried out to investigate project planning and control systems and their functionalities. Through meticulous analysis of the content of 204 papers, a total of 36 single and integrated planning methods, 82 control metrics, and 10 collaborative planning methods were subject to comprehensive examination. The objectives and aims inherent to these planning and control systems were defined as their respective functionalities. The processes of extracting and analyzing functionalities are detailed in studies by Sheikhkhoshkar et al. (2023b), Sheikhkhoshkar et al. (2023a), Sheikhkhoshkar et al. (2024b), and Sheikhkhoshkar et al. (2024a). This phase of the research yielded a total of 148 functionalities as input data for further analysis, as illustrated in Figure 2 and drafted in Table A1 within Appendix A. It is noteworthy that all functionalities were extracted verbatim, maintaining the exact phrasing as presented in each respective paper, without altering their structure.

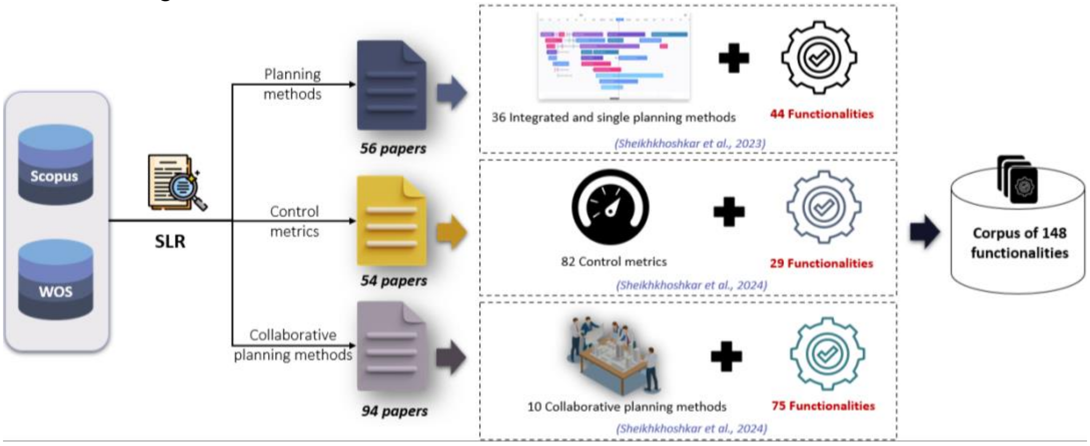


Figure 2: An outline of the procedure for extracting functionalities.

3.2 Data Analysis

This research has adopted advanced text mining analysis as the main approach to analyze data and meet the study's goals. Text mining involves examining text to extract useful information for various purposes (Zhou et al. 2019). A variety of text mining techniques have been employed, utilizing the Orange data mining tool (Demšar et al. 2004), which allows workflows to be designed and created by connecting predefined or user-designed components called widgets. This part examines the sequential steps employed for the data analysis, as outlined in Figure 3. The data analysis phase encompasses several pivotal processes, including data preprocessing, vectorization and text embedding using SBERT, clustering analysis, and the generation of word clouds alongside keyword extraction. The following sections delve into the specifics of each step.

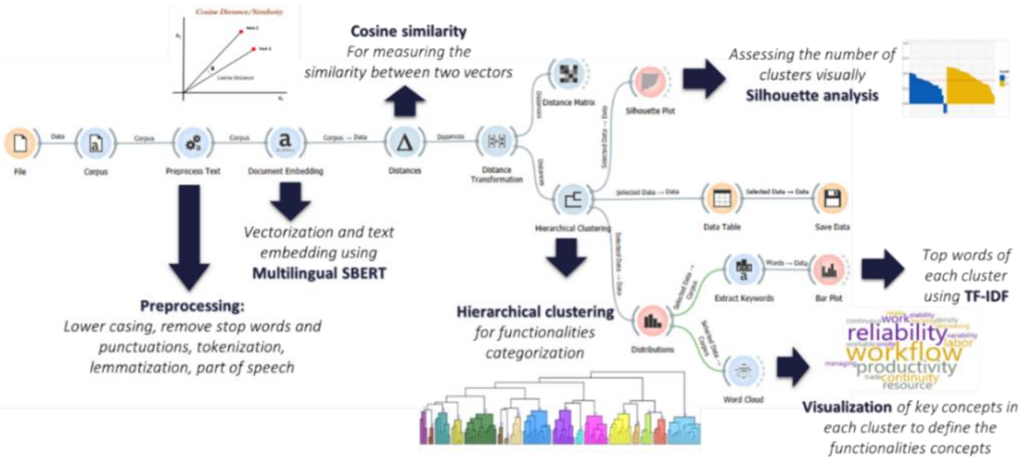


Figure 3: The employed steps for data analysis.

Adopting this complex and quantitative approach offers significant advantages over purely qualitative analysis and classification. While qualitative methods provide valuable insights and deep understanding through subjective interpretation, they often lack the scalability needed for comprehensive analysis across large datasets (Jacobs and Tschötschel 2019). Therefore, automated clustering approaches are more scalable and efficient than manual qualitative analysis for a large number of phrases. Integration of the semantic textual similarity and clustering methods allows a more objective and data-driven analysis of the 148 phrases (functionalities). These techniques minimize the potential for subjective bias that can sometimes affect qualitative analysis. Moreover, advanced computational methods in such studies can uncover patterns and relationships that may not be immediately apparent through qualitative methods (Abdullah et al. 2023). Clustering, in particular, can reveal natural groupings within the data, offering a deeper understanding of the underlying concepts.

The results from these quantitative methods provide a robust foundation for subsequent qualitative analysis. Once the initial clusters and underlying concepts are identified, they can be further explored and validated through qualitative methods such as expert interviews, offering a comprehensive mixed-method approach. By integrating semantic textual similarity and clustering with potential subsequent qualitative analysis, we aim to provide a balanced, thorough, and methodologically sound exploration of the functionalities in project planning and control systems. This approach leverages the strengths of both quantitative and qualitative methodologies to achieve a holistic understanding.

3.2.1 Data Preprocessing

To prepare the raw text data for further processing and analysis, several preprocessing steps were executed on the extracted functionalities. These steps include converting all letters to lowercase, removing stop words and punctuation, tokenization, lemmatization, and part-of-speech tagging. Lowercasing normalizes the text by equating variations in capitalization, thus providing a uniform basis for analysis. The elimination of stop words and punctuation concentrates the analysis on more substantive words and purifies the textual data. Tokenization segments the text into discrete units, or tokens, which are crucial for subsequent processing steps. Lemmatization simplifies words to their lemma forms, enhancing the model’s capacity to generalize by focusing on the fundamental meaning of words across various morphological manifestations. Part of speech tagging adds syntactic information, aiding in understanding the grammatical context of words. These preprocessing steps are pivotal in transforming raw text into a structured format, which is critical for various NLP applications such as semantic textual similarity. Table 1 presents the applied preprocessing steps to a functionality.

Table 1: Applied preprocessing steps to a sample functionality.

Original extracted functionality	Preprocessing steps	Preprocessed functionality
Visualization of the schedules to understand/communicate content to a variety of stakeholders	Lowercasing	*visualization of the schedules to understand/ communicate content to a variety of stakeholders*
	Removing stop words	*visualization schedules understand/ communicate content variety stakeholders*
	Removing punctuation	*visualization schedules understand communicate content variety stakeholders*
	Tokenization	*visualization, schedules, understand, communicate, content, variety, stakeholders*
	Lemmatization	*visualization, schedule, understand, communicate, content, variety, stakeholder*
	Part of speech tags	*visualization/NN, schedules/NNS, understand/VBP, communicate/NN, content/NN, variety/NN, stakeholders/NNS*

3.2.2 Semantic Textual Similarity (STS)

To capture the semantic similarity between functionalities, the Sentence-BERT (SBERT) neural network model was employed, which generates vector representations for sentences and phrases. The BERT model is a deep learning algorithm based on the transformer architecture that is designed to understand the context of language in text. It utilizes a mechanism known as attention, which weighs the relative importance of words in a phrase. BERT models are pre-trained on a large corpus of text and then fine-tuned for specific tasks (Moon et al. 2022). While BERT is highly effective at understanding context and meaning, the original model was not optimized for generating sentence-level embeddings. To address this, the Sentence-BERT (SBERT) model was developed (Reimers and Gurevych 2019). SBERT is a modification of the pre-trained BERT network that uses Siamese and triplet network structures to derive semantically meaningful embeddings that are well-suited for clustering, semantic search, and other similarity tasks (Reimers and Gurevych 2019). SBERT modifies the final embedding process to produce fixed-length vector representations of entire sentences rather than individual tokens. The SBERT architecture comprises the following elements:

1- ***Tokenizer:*** It first tokenizes sentences into a series of word pieces

2- ***BERT encoders:*** are composed of several layers of transformer blocks that process the tokenized input, applying self-attention and feed-forward neural network operations to encode contextual information into vector form.

3- ***Pooling layer:*** SBERT applies a pooling strategy to the output of the encoders to derive a single fixed-size sentence embedding.

4- ***Fine-Tuning:*** SBERT is fine-tuned on Natural Language Inference (NLI) tasks, which helps the embeddings to capture sentence-level semantic relationships effectively.

3.2.3 Clustering Analysis

Following vector embeddings by the SBERT model, the cosine similarity was used to evaluate the semantic similarity between functionalities. It computes the cosine of the angle between two vectors in a multi-dimensional space, with values ranging from -1 (dissimilar) to 1 (very similar). Based on this metric, hierarchical clustering was applied to classify the functionalities into distinct clusters. Hierarchical clustering is an unsupervised machine-learning technique that groups entities by their similarity or distance to form a multi-level hierarchy of clusters (Reddy and Vinzamuri 2018). The success of hierarchical clustering significantly depends on the considered linkage method, which determines the metric for calculating distances between clusters (Contreras and Murtagh 2015). The linkage methods, including single, complete, average, and ward linkage, offer different approaches to defining inter-cluster similarity. Throughout the clustering process, various linkage methods were iteratively employed, with Silhouette analysis utilized at each iteration to assess clustering performance. Subsequently, ward linkage emerged as the preferred method due to its ability to generate well-defined and coherent clusters.

The hierarchical clustering technique was particularly suitable for this study as it enabled the establishment of a multi-level taxonomy of sentence groupings, which reflects the fine spectrum of semantic relationships within our corpus. The dendrogram structure derived from hierarchical clustering furnished a visual depiction of the functionalities clusters, thereby aiding in the intuitive understanding of the diverse categories inherent in the functionalities.

To validate the results of hierarchical clustering and ensure that functionalities with similar meanings were appropriately grouped, a Silhouette analysis was performed complemented by the insights of domain experts. The Silhouette analysis provided a quantitative measure of how well each sentence fit within its cluster, which was crucial for assessing the cohesion and separation of the clusters (Abdul Nabi and El-adaway 2022). In parallel, semi-structured interviews and surveys were conducted to seek the qualitative judgment of domain experts. To do so, using a purposive sampling approach, experts were selected based on three key factors: 1) their practical experience and theoretical understanding of the study's topics, including traditional and lean-based planning and scheduling methods, control systems, and collaborative planning methods; 2) diversity in their professional roles, incorporating both academic and industry experts with different relevant roles to ensure various skills and perspectives; and 3) their willingness to actively engage in feedback sessions. A targeted cohort of 18 individuals was identified by reviewing LinkedIn profiles, the IsoBIM project's partners, and various online repositories. Invitations to participate were then sent via email and direct messaging channels. Of these, ten individuals responded affirmatively and were subsequently scheduled for interviews. This sample size was sufficient, as many studies employing purposive sampling opt for sample sizes ranging from 5 to 25 (Jepson et al. 2020; Zulu et al. 2023). The selected experts' qualifications are outlined in Table 2. The interview sessions constituted a component of a six-month endeavor involving the design and evaluation of a comprehensive multi-level and collaborative project planning and control framework within the IsoBIM project which outlines a collaborative approach for renovating buildings with external insulation based on lean and BIM paradigms.

This allowed the incorporation of their knowledge and expertise in the evaluation process, creating a robust mechanism for verifying the semantic integrity of the clustering approach. The combination of these methods ensured a thorough validation of the clustering, confirming that similar functionalities were indeed categorized together.

Table 2: Experts' profile.

Type of experts	Experts code	Background/role	Experience in project planning and control (years)
Industry experts	IE1	Senior program scheduler	20-25
	IE2	Senior project manager	15-20
	IE3	Consultant lean management	10-15
	IE4	Senior project construction manager	15-20
	IE5	CEO of a lean construction software company	20-25
Academic experts	AE1	Senior lecturer in the built environment	15-20
	AE2	Full professor in construction management	25-30
	AE3	Assistant professor in construction management	5-10
	AE4	Associate professor in production planning and control	15-20
	AE5	Full professor in production management and control	15-20

3.2.4 Concept Identification and Taxonomy Building

Word clouds and keyword extractions for each cluster were created after the functionalities had been clustered, serving as tools for concept identification. These visualizations aid in simplifying each cluster by highlighting the most prevalent terms and facilitating an intuitive grasp of the underlying concepts. Subsequently, domain experts were engaged through semi-structured interviews and surveys to carefully examine and match the identified concepts to the schedule levels. Their invaluable insights allowed us to organize these concepts into a structured hierarchy, laying the foundation for a comprehensive taxonomy of project planning and control systems' functionalities.

4. RESULTS AND ANALYSIS

This section presents and discusses the research findings.

4.1 Clustering and Concept Identification Results

The dendrogram plot of hierarchical clustering, depicted in Figure 4, outlines the hierarchical arrangement of 17 distinct clusters devised to categorize 148 functionalities. A Silhouette plot was used to assess cluster integrity. The silhouette coefficient ranges from -1 to 1, where a high value indicates that the object is well-matched to its own cluster and poorly matched to neighbouring clusters. The analysis indicates that the majority of functionalities possess positive Silhouette scores, which point to well-defined clusters. Nonetheless, a few clusters display bars nearing zero or dipping into negative values, signalling a potential overlap between clusters or a less robust clustering configuration for certain functionalities. This indicates areas where cluster structure might be improved. These findings underscore the importance of continuous refinement in the clustering process to ensure precise and meaningful categorization.

Furthermore, Table 3 provides more details on the clustering outcomes, encompassing details such as the number of functionalities, top words, silhouette plots, word clouds, and pivotal concepts associated with each cluster. Notably, the clusters exhibited variance in size and thematic focus, with the count of functionalities per cluster ranging from 3 to 17. The identified key concepts within these clusters spanned a wide spectrum of project management activities crucial for planning and control, like collaboration and communication management and data-driven decision-making. The importance of specific terms and concepts within each cluster was visualized through word clouds. As illustrated in Table 3, most clusters, such as clusters 4, 5, 7, 10, 11, 12, 13, 14, 15, 16, and 17 exhibit a singular key concept, whereas in other clusters, multiple concepts are identified. For instance, within cluster 6, the concepts of productivity, reliability, and workflow management were observed concurrently.

The silhouette plots provide a quantitative evaluation of the fit for each functionality within its cluster. Some clusters, including 5, 9, 13, and 15, exhibited less cohesion than anticipated. This suboptimal clustering may be ascribed to the inherent characteristics of the SBERT model, which is principally optimized for sentence-level inputs rather than phrases. The input data for this study consisted of phrases, which typically do not encapsulate the full semantic breadth that complete sentences offer, posing a challenge for effective embedding by SBERT. Moreover, due to the absence of a robust and comprehensive dataset characterizing the functionalities of project planning and control systems, essential for the training and fine-tuning of domain-specific models, this study adopted a pre-trained SBERT model. Such pre-trained models may encounter constraints when attempting to effectively generalize to domain-specific textual data (Wang et al. 2021). This may constitute an additional rationale for the presence of multiple concepts within certain clusters.

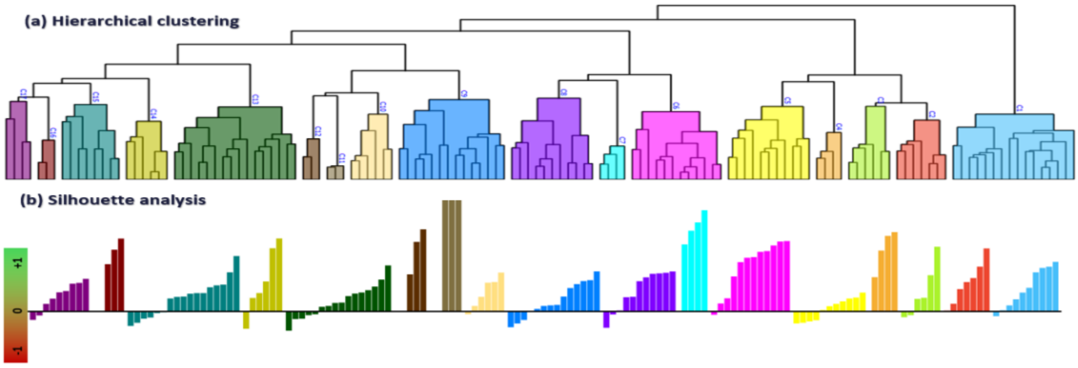


Figure 4: Hierarchical clustering results.

Table 3: More details of clustering and concept identification outcomes.

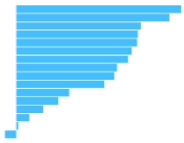
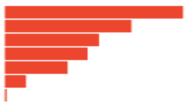
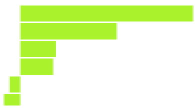
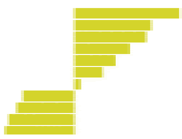
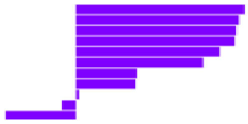
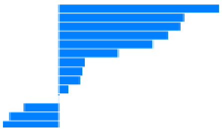
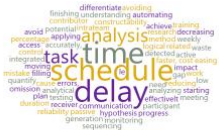
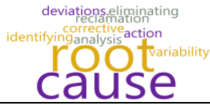
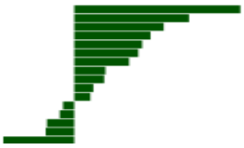
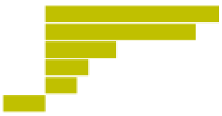
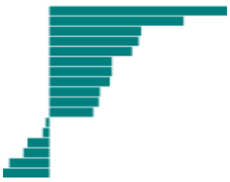
Cluster	No. of functionalities	Top words	Silhouette plot -1 0 +1	Word cloud	Key concept
1	16	Collaboration, project, stakeholder, team, communication			**Collaboration and communication management**
2	7	Communication, real-time, worker, site			**Communication management and real-time tracking**
3	7	Site, safety, logistics, layout			**Safety and logistics management**
4	4	Conflict, solving, collision, detection			**Conflict management**
5	11	Visualization, project, sequence, real-time			**Visualization**
6	12	Reliability, productivity, workflow, continuity			**Productivity, reliability and workflow management**
7	4	Supply, chain, process, instability			**Supply chain management**

Table 3: (Continued).

Cluster	No. of functionalities	Top words	Silhouette plot (-1 0 +1)	Word cloud	Key concept
8	**11**	**Resource, waste, information, production**			**Resource and waste management**
9	**14**	**Schedule, delay, time, analysis, task**			**Delay and contract management**
10	**6**	**Constraint, pull, systematic, removal**			**Constraint management**
11	**3**	**Uncertainty, risk, managing**			**Uncertainty and risk management**
12	3	Root, cause, analysis, corrective, action			**Root cause analysis**
13	16	Construction, planning, integration, alignment			**Integration management**
14	6	Continuous, learning, improvement, lesson, learn			**Learning and knowledge sharing**
15	17	Project, performance, control, progress, monitoring			**Project performance management**
16	3	Commitment, reliability, plan			**Commitment management**
17	9	Managerial, decision, information, making, process			**Information-driven decision making**

To deal with these challenges, an evaluation by domain experts was deemed necessary. To do so, Based on the results of clustering and identified concepts, clusters 1, 2, 6, and 8, which initially contained two or more concepts, were modified so that each cluster contained a single concept. Consequently, a total of 20 clusters were considered, each containing only one key concept to validate the identified concepts within each cluster using experts' insights. This decision was made to ensure a more granular and thorough analysis, allowing for identifying and validating specific concepts that might have been grouped initially due to the pre-trained SBERT model's domain-specific limitations.

Semi-structured interviews and surveys were utilized to capture the experts' specialized insights into the meaning and pertinence of functionalities within the clusters. During the interviews, the participants were asked about the number of functionalities in each cluster that matched the cluster's main concept and their meaningful grouping. They were also queried concerning the functionalities inside a cluster that did not follow the same concept as others. Table 4 presents the outcomes obtained from expert evaluations regarding how well functionalities align with each cluster's core concept.

The findings reveal that, with an average 87% alignment rate, the functionalities received consistency with their respective primary concepts. Further elaboration on the alignment of functionalities within each cluster is provided in Table 4. Despite the refinement of clusters guided by experts' feedback, the identified key concepts for constructing the taxonomy remained consistent with those established through hierarchical clustering. As a result, 20 key concepts were identified and considered for building the taxonomy of project planning and control systems' functionalities. The identified functional concepts serve dual roles within the context of project planning and control systems. From one angle, these concepts highlight the complexity and underlying managerial facets inherent to these systems. From another perspective, they function as essential requirements for the design and implementation of project planning and control systems. By mapping both single and integrated planning methods and control systems with these identified concepts, a deeper level of understanding and sense-making will be fostered among project stakeholders regarding these methods and systems. This will pave the way for future research endeavours in this domain. Consequently, this enhanced comprehension enables stakeholders to make more informed decisions when selecting the most suitable planning and control approach based on the specific requirements of their projects.

Table 4: Outcomes derived from experts' evaluation.

Functionality main concepts	**No. of items**	**Experts**										**Average alignment rate**
		1	2	3	4	5	6	7	8	9	10	
Collaboration management	8	100%	100%	88%	88%	100%	100%	88%	100%	100%	100%	96%
Commitment management	5	80%	80%	100%	100%	80%	80%	100%	80%	80%	80%	86%
Communication management	11	64%	73%	82%	82%	91%	82%	73%	73%	82%	73%	77%
Conflict management	4	100%	100%	100%	75%	100%	100%	100%	75%	75%	75%	90%
Constraint management	6	83%	67%	67%	83%	83%	83%	83%	67%	67%	83%	77%
Contract and delay management	13	77%	85%	77%	69%	77%	77%	69%	85%	77%	85%	78%
Integration management	14	57%	57%	64%	71%	64%	79%	71%	64%	64%	57%	65%
Learning and knowledge sharing	6	100%	100%	83%	83%	83%	100%	83%	83%	100%	100%	92%
Workflow management	7	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
Project performance management	17	94%	88%	100%	100%	94%	94%	88%	88%	100%	100%	95%
Real-time monitoring	4	100%	100%	100%	75%	75%	100%	100%	100%	75%	100%	93%
Reliability management	5	60%	80%	60%	60%	80%	80%	60%	80%	60%	80%	70%
Resources management	7	86%	86%	71%	86%	86%	71%	71%	71%	86%	86%	80%
Root cause analysis	3	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
Safety and logistics management	4	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
Supply chain management	4	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
Uncertainty and risk management	4	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
Visualization	11	73%	73%	82%	82%	73%	91%	82%	73%	82%	91%	80%
Waste management	6	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
Information-driven decision making	9	67%	67%	78%	78%	78%	67%	67%	56%	78%	67%	70%
											Total average alignment rate	**87%**

In contrast to the conventional perspective regarding the primary objectives of project planning and control systems, which typically emphasize time and cost management, this research has illustrated that these systems encompass a variety of functions and concepts, including collaboration management, communication management, conflict management, resource management, safety and logistics management, among others. As such, it becomes imperative for project managers and their teams to initially find out their functional requirements based on these identified concepts. They can then endeavour to identify the most suitable planning and control system tailored to meet these requirements. These insights not only highlight the multifaceted nature of project planning and control systems but also underscore the significance of considering diverse managerial aspects and functions. Such considerations are instrumental in empowering project teams to make more informed and effective decisions regarding their planning and control systems.

After the validation of functionality concepts, to construct the taxonomy, a survey was designed and shared with selected experts during another round of semi-structured interviews. The experts were asked to assess the significance of each functionality concept across various schedule levels using a Likert scale ranging from 0 (no importance) to 4 (very high importance). Following that, a taxonomy for project planning and control systems' functionalities was developed based on the experts' average rating of the relative importance of each functionality concept across schedule levels (Figure 5).

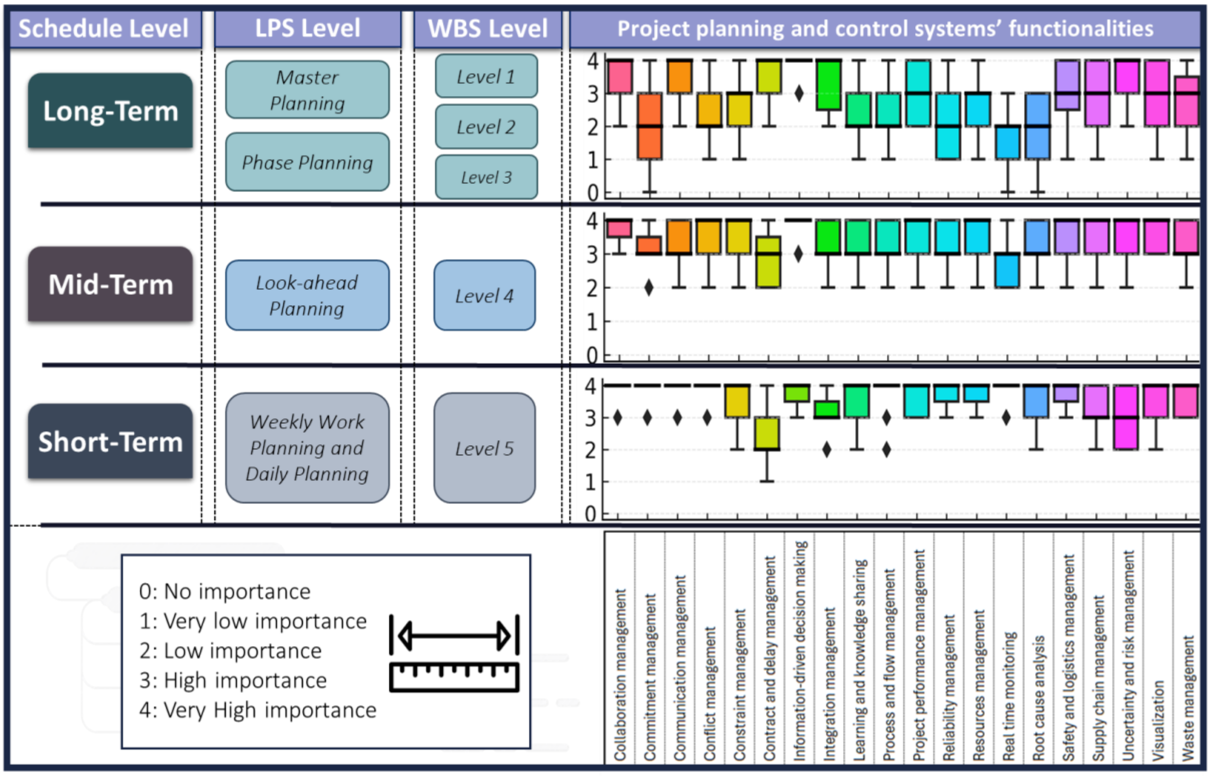


Figure 5: Taxonomy of multi-level project planning and control systems' functionalities.

Figure 5 represents the importance of various functions across different scheduling levels. As can be seen, certain functions are deemed more critical than others, denoted by the importance ratings (2 for low importance, 3 for high importance, and 4 for very high importance). Based on the standard deviations observed in the box plots, the results generally exhibit high reliability due to the low variability in expert evaluations for each functionality across short-term and mid-term schedule levels. However, for the long-term level, there is a slight increase in deviation, suggesting that opinions vary more regarding the importance of some functionalities in the long term. Despite this, the overall consistency across all schedule levels indicates a broad consensus among experts, making the assessments reliable for decision-making and strategic planning.

Information-driven decision-making is highlighted as having very high importance across all scheduling levels. This emphasis across long-term, mid-term, and short-term schedules underscores the critical role of accurate and timely information in guiding all stages of project management. At the long-term level, data-driven decisions are fundamental to establishing project objectives and aligning them with strategic goals. In mid-term scheduling, such decision-making is vital to adapting the project plan based on progress and external changes, ensuring the project remains on track. At the short-term level, it is necessary for daily operations, where immediate decisions can have significant impacts on the day-to-day success of project tasks.

In the mid-term schedule level, supply chain management, uncertainty and risk management, and collaboration management are particularly emphasized. Supply chain management is crucial in this phase as it ensures the timely availability of materials and resources, which are essential for maintaining project momentum. Uncertainty and risk management are also critical since this is the stage where potential risks must be identified and mitigated to prevent project delays. Collaboration management gains significance due to the increasing need for coordinating multiple teams and stakeholders as the project moves from planning to execution. Furthermore, at this level, constraint management, conflict resolution, resource allocation, and visualization are recognized for their pivotal roles in project planning and control. Effective constraint management facilitates teams in anticipating and navigating project limitations and bottlenecks to ensure the achievement of project objectives. Conflict resolution ensures smooth collaboration by addressing team disputes and identifying and resolving temporal or spatial conflicts to maintain project continuity. To mitigate delays and enhance operational efficiency, it is crucial to allocate resources as effectively as possible to optimize capacities and make effective use of materials and equipment. Moreover, visualization facilitates transparent communication, which enables stakeholders to understand various what-if scenarios and construction sequences to make well-informed decisions promptly. Collectively, these elements provide the foundation for successful project execution, aligning team endeavours with project objectives.

At the short-term level, most functions receive very high importance based on expert feedback, which can be attributed to the immediacy and tactical nature of this phase. Effective execution at this stage is essential for the daily progress and ultimate success of the project. This phase is where planning is translated into action, and as such, all functions, from resource allocation to safety and logistics, must be managed with a high degree of precision and responsiveness. The feedback from experts likely reflects the reality that, at this level, there is a smaller margin for error, and the impact of each function is immediately observable on the project's progress.

By systematically classifying and organizing identified functionality concepts according to their significance at each scheduling level, it becomes feasible to align the features and functions of various planning methods with the corresponding functionalities at each scheduling level, thereby facilitating the selection of the most effective approach for managing that specific level. For instance, at the mid-term schedule level, where collaboration management, constraint management, supply chain management, and workflow management are prioritized, the features and functionalities of planning systems such as the last planner system and takt time planning, particularly in managing lookahead plans, are crucial. These systems can significantly influence the management of these aspects, enhancing overall project coordination and efficiency.

This taxonomy acts as a scaffold for future research, providing a structured framework to explore the interplay between different planning and control systems and their respective functionalities and project teams' functional requirements at each scheduling level. It invites scholars and practitioners to consider project planning and control not merely as a collection of disparate methods but as a cohesive system that operates across multiple levels, each with its distinct functions and requirements.

By identifying the specific functionalities pertinent to each level of scheduling, the proposed taxonomy paves the way for developing a multi-level project planning and control system that integrates various functions and methods to address the unique demands of each scheduling level. The approach advances the field by encouraging a systemic view of project planning and control, where the synergy between different levels and functions can lead to more robust and adaptive management practices. Consequently, this perspective enhances the potential for achieving strategic alignment and operational efficiency, leading to improved project performance and successful outcomes.

5. DISCUSSION

This research adopts a comprehensive and multi-stage methodology to investigate the functionalities of project planning and control systems in the construction industry. Initially, a thorough data collection process is conducted through systematic literature reviews and meta-analyses, focusing on various planning methods and control metrics. This phase extracts 148 functionalities, serving as a foundational dataset for deeper analysis. The extracted data undergo a series of preprocessing steps, including lowercasing, removal of stop words and punctuation, tokenization, lemmatization, and part-of-speech tagging. These preprocessing efforts are crucial for standardizing the data, thus facilitating more sophisticated text-mining techniques.

Semantic analysis plays a pivotal role in the methodology, where the Sentence-BERT (SBERT) model is employed to generate text embeddings. This model captures the semantic similarities between the functionalities, enabling the application of hierarchical clustering to group functionalities into semantically coherent clusters. This clustering is crucial for developing a structured taxonomy of functionalities, reflecting the semantic relationships within the data.

The results from the clustering process are rigorously analyzed, with the study achieving 17 distinct clusters of functionalities. The integrity and appropriateness of these clusters are quantitatively assessed using Silhouette analysis, which generally indicates a strong alignment of functionalities within their respective clusters. However, some clusters exhibit potential overlaps or less distinct boundaries, suggesting areas for further refinement. The reason could be that this study utilized a pre-trained SBERT model due to the lack of a thorough dataset on functionalities of project planning and control systems. While pre-trained models like SBERT are useful, they may struggle to generalize effectively to domain-specific data, potentially leading to multiple concepts within certain clusters.

The validation of the clustering results involves a comprehensive approach, employing word clouds, keyword extraction, and expert interviews. These methods ensure that the clusters are not only statistically valid but also practically relevant and intuitive. Experts from the industry and academia are engaged to assess the alignment of functionalities within clusters, with findings showing an average 87% alignment rate. This high rate of consistency confirms the effectiveness of the semantic clustering approach and underscores the robustness of the taxonomy developed.

Finally, the construction of a functional taxonomy discusses based on the importance of each functionality concept across different scheduling levels, considering experts' insight. This taxonomy is instrumental for practitioners, allowing project managers to select and implement planning and control systems that are best suited to the specific needs of their projects at various stages. By systematically classifying and organizing functionality concepts, the research provides a valuable framework that aids in understanding the multifaceted nature of project planning and control systems.

The importance of functionalities, as reflected in the standard deviations of experts' feedback, shows high reliability due to the generally low variability in expert assessments for each functionality across various schedule levels. This consistency across all schedule levels points to broad agreement among experts, making these evaluations dependable for decision-making and strategic planning.

The identified functionality concepts extracted from various planning methods and control metrics as a common concept across these systems can play a role as a primary key concept in the database management domain. This unique position allows them to integrate planning methods and control metrics across different scheduling levels, paving the way for developing a multi-level and integrated planning and control system. Such an approach can effectively address the limitations inherent in individual planning methods and control systems by leveraging and integrating the advantages and functionalities of each. In forthcoming research, efforts will be concentrated on exploring this integration further. The established taxonomy and functionality concepts will form the foundational basis for developing a comprehensive, multi-level, and integrated planning and control system. This innovative approach aims to enhance the robustness and adaptability of project management strategies, ensuring that they are well-suited to meet the diverse and dynamic needs of construction projects.

In essence, this study exemplifies a methodologically rigorous approach to dissecting and categorizing project planning functionalities, highlighting the importance of combining quantitative methods with qualitative insights to develop tools that are not only theoretically sound but also practically applicable in improving project management practices.

6. THE IMPLICATIONS OF THE FINDINGS

The findings from this research offer profound scientific and practical implications in the project planning and control body of knowledge. Scientifically, the research introduces a novel and robust multi-step methodology that combines advanced text mining techniques with expert validation, setting a new benchmark for future studies in project management and related disciplines. This approach allows for a systematic understanding and integration of various project planning methods and control metrics, enhancing the theoretical framework within which these systems are analyzed and applied.

Practically, the identified functionality concepts and developed taxonomy from this research provide actionable benefits for project managers and stakeholders. Given the often low level of knowledge about planning and control systems among stakeholders, it is challenging for them to select an appropriate planning and control system for their projects. To deal with this issue, the identified functionality concepts in this study can act as the functional requirement of project planning and control systems that will aid the project team to more effectively determine their requirements and then see what planning and control system is more aligned with those requirements. Additionally, the developed taxonomy will help the project team understand how much effort should be put into each level of schedule for each functionality, optimizing efforts and resources. Furthermore, the identified functionality concepts and developed taxonomy can be used as a foundation for extracting and formulizing the tactic knowledge of the domain expert and building a knowledge repository of project planning and control systems. This knowledge repository can act as a database in an expert system to suggest the most-fitted planning and control systems for the construction project, considering the project team's requirements.

7. CONCLUSION, LIMITATIONS, AND FUTURE DIRECTIONS

This study developed a robust taxonomy of functionalities within project planning and control systems through a rigorous methodology. A wide range of planning and control systems' functionalities were identified and classified into distinct and semantically coherent clusters by combining systematic literature reviews with advanced text mining techniques, including sentence-BERT for semantic analysis and hierarchical clustering. Incorporating both a quantitative approach and qualitative analysis through expert feedback in validating the clusters and concepts ensures that the taxonomy reflects the realities of construction project management. This validation confirms the applicability and significance of the research findings to the project planning and control domain. The outcomes of this study furnish a structured framework for comprehending the diverse functionalities inherent in project planning and control systems. It encapsulates the complexities and underlying concepts of various planning and control methods, which enhances actionable insights for project managers and stakeholders.

The proposed taxonomy makes a substantial contribution to the field of project planning and control by mapping structured functionality concepts across different schedule levels against planning methods and control metrics. This facilitates project managers in selecting appropriate planning methods tailored to address specific project requirements, potentially leading to enhanced decision-making, superior project outcomes, and improved alignment between project objectives and execution strategies. The scientific implications of this research are profound, setting new processes for integrating quantitative and qualitative data analysis techniques in project management studies. The practical ramifications are equally significant, providing project managers with a deeper understanding of the multifaceted nature of project planning and control systems and enabling the use of this taxonomy to improve decision-making processes.

A notable limitation of this research was the absence of a domain-specific textual dataset, which necessitated the use of a pre-trained SBERT model. While SBERT is proficient in processing and analyzing text, it was not originally optimized for semantic textual similarity analysis specifically within the construction domain. This limitation underscores the potential discrepancies in semantic understanding when applying generalized models to specialized fields.

Looking ahead, future research should focus on developing customized models that are directly trained on domain-specific datasets within the construction industry. Moreover, future endeavours may find value in utilizing the derived taxonomy and functional concepts outlined in this study as a foundation for assessing the efficacy of various planning methods and control metrics in supporting these functionality concepts. Such an approach could propose a multi-level planning and control system that accounts for the diverse functional requirements of project teams. Eventually, this study not only contributes to the academic literature but also offers tangible strategies for promoting project planning and control in the construction sector by bridging the gap between theoretical research and practical application.

ACKNOWLEDGEMENTS

The study was conducted at the Nancy Automation Research Center (CRAN) and received financial support from the National Research Agency of France as part of the IsoBIM project (grant number ANR-20-CE10-0011).

REFERENCES

Abdul Nabi, M., and El-adaway, I. H. (2022). "Understanding disputes in modular construction projects: Key common causes and their associations." Journal of Construction Engineering and Management, 148(1), 04021184, https://doi.org/10.1061/(ASCE)CO.1943-7862.0002208.

Abdullah, M. H. A., Aziz, N., Abdulkadir, S. J., Alhussian, H. S. A., and Talpur, N. (2023). "Systematic literature review of information extraction from textual data: recent methods, applications, trends, and challenges." IEEE Access, 11, 10535-10562, https://doi.org/10.1109/ACCESS.2023.3240898.

Al Qady, M., and Kandil, A. (2010). "Concept relation extraction from construction documents using natural language processing." Journal of construction engineering and management, 136(3), 294-302, https://doi.org/10.1061/(ASCE)CO.1943-7862.0002172.

AlNasseri, H., and Aulin, R. (2015). "Assessing understanding of planning and scheduling theory and practice on construction projects." Engineering Management Journal, 27(2), 58-72, https://doi.org/10.1080/10429247.2015.1035963.

Arayici, Y., Tokdemir, O. B., and Kassem, M. (2023). "A quantitative, evidence-based analysis of correlations between lean construction and building information modelling." Smart and Sustainable Built Environment, 12(5), 975-1001, https://doi.org/10.1108/SASBE-03-2022-0052

Ballard, G., and Tommelein, I. (2021). "2020 Current process benchmark for the last planner (R) system of project planning and control." https://doi.org/10.34942/P2F593.

Contreras, P., and Murtagh, F. (2015). "Hierarchical clustering." Handbook of cluster analysis, 103-123, https://doi.org/10.1201/b19706.

Demšar, J., Zupan, B., Leban, G., and Curk, T. (2004). "Orange: From experimental machine learning to interactive data mining." Proc., Knowledge Discovery in Databases: PKDD 2004: 8th European Conference on Principles and Practice of Knowledge Discovery in Databases, Pisa, Italy, September 20-24, 2004. Proceedings 8, Springer, 537-539, 3540231080, https://doi.org/10.1007/978-3-540-30116-5_58.

Ezzeddine, A., Shehab, L., Lucko, G., and Hamzeh, F. (2022). "Forecasting construction project performance with momentum using singularity functions in LPS." Journal of Construction Engineering and Management, 148(8), 04022063, https://doi.org/10.1061/(ASCE)CO.1943-7862.0002320.

Frandson, A. (2019). Takt time planning as a work structuring method to improve construction work flow, University of California, Berkeley, https://escholarship.org/uc/item/6dp4n4fz.

Ghanem, M., Hamzeh, F., Seppänen, O., Shehab, L., and Zankoul, E. (2022). "Pull planning versus push planning: Investigating impacts on crew performance from a location-based perspective." Frontiers in Built Environment, 8, 980023, https://doi.org/10.3389/fbuil.2022.980023.

Hamzeh, F. R., El Samad, G., and Emdanat, S. (2019). "Advanced metrics for construction planning." Journal of Construction Engineering and Management, 145(11), 04019063, https://doi.org/10.1061/(ASCE)CO.1943-7862.0001702.

Hu, X., Cui, N., Demeulemeester, E., and Bie, L. (2016). "Incorporation of activity sensitivity measures into buffer management to manage project schedule risk." European Journal of Operational Research, 249(2), 717-727, https://doi.org/10.1016/j.ejor.2015.08.066.

Jacobs, T., and Tschötschel, R. (2019). "Topic models meet discourse analysis: a quantitative tool for a qualitative approach." International Journal of Social Research Methodology, 22(5), 469-485, https://doi.org/10.1080/13645579.2019.1576317.

Jepson, J., Kirytopoulos, K., and London, K. (2020). "Insights into the application of risk tools and techniques by construction project managers." International Journal of Construction Management, 20(8), 848-866, https://doi.org/10.1080/15623599.2018.1494673.

Jung, Y., and Golparvar-Fard, M. (2023). "EXPLORING THE USE OF NLP TO AUTO-ALIGN MASTER SCHEDULES WITH SUPERINTENDENT'S LOOK-AHEADS IN CONSTRUCTION PROJECTS." Proceedings of International Structural Engineering and Construction, 10(1), CON-30-31-CON-30-36, www.doi.org/10.14455/ISEC.2023.10(1).CON-30.

Jung, Y., Hockenmaier, J., and Golparvar-Fard, M. (2024). "Transformer language model for mapping construction schedule activities to uniformat categories." Automation in Construction, 157, 105183, https://doi.org/10.1016/j.autcon.2023.105183.

Kang, Y., Cai, Z., Tan, C.-W., Huang, Q., and Liu, H. (2020). "Natural language processing (NLP) in management research: A literature review." Journal of Management Analytics, 7(2), 139-172, https://doi.org/10.1080/23270012.2020.1756939.

Ko, T., David Jeong, H., and Lee, J. (2023). "Natural language processing–driven similar project determination using project scope statements." Journal of Management in Engineering, 39(3), 04023005, https://doi.org/10.1061/JMENEA.MEENG-5229.

Ko, T., Jeong, H. D., and Lee, G. (2021). "Natural language processing–driven model to extract contract change reasons and altered work items for advanced retrieval of change orders." Journal of Construction Engineering and Management, 147(11), 04021147, https://doi.org/10.1061/(ASCE)CO.1943-7862.0002172.

Laufer, A., and Tucker, R. L. (1987). "Is construction project planning really doing its job? A critical examination of focus, role and process." Construction management and economics, 5(3), 243-266, https://doi.org/10.1080/01446198700000023.

Li, S., Wang, J., and Xu, Z. (2024). "Automated compliance checking for BIM models based on Chinese-NLP and knowledge graph: an integrative conceptual framework." Engineering, Construction and Architectural Management, https://doi.org/10.1108/ECAM-10-2023-1037.

Moon, S., Chi, S., and Im, S.-B. (2022). "Automated detection of contractual risk clauses from construction specifications using bidirectional encoder representations from transformers (BERT)." Automation in Construction, 142, 104465, https://doi.org/10.1016/j.autcon.2022.104465.

Naderi, M., Nazari, A., Shafaat, A., and Abrishami, S. (2024). "Enhancing accuracy in construction overhead cost estimation: a novel integration of activity-based costing and building information modelling." Smart and Sustainable Built Environment, https://doi.org/10.1108/SASBE-07-2023-0180.

Nwadigo, O. B.-K., Naismith, N., GhaffarianHoseini, A., GhaffarianHoseini, A., and Tookey, J. (2021). "Construction project planning and scheduling as a dynamic system: a content analysis of the current status, technologies and forward action." Smart and Sustainable Built Environment, 11(4), 972-995, https://doi.org/10.1108/SASBE-02-2021-0022.

O. AlSehaimi, A., Tzortzopoulos Fazenda, P., and Koskela, L. (2014). "Improving construction management practice with the Last Planner System: a case study." Engineering, Construction and Architectural Management, 21(1), 51-64, https://doi.org/10.1108/ECAM-03-2012-0032.

Olivieri, H., Seppänen, O., and Granja, A. D. (2016). "Integrating Lbms, Lps and Cpm: a practical process." Proc., Annual Conference of the International Group for Lean Construction, National Pingtung University of Science and Technology, 3-12, http://iglc2016.com/.

Reddy, C. K., and Vinzamuri, B. (2018). "A survey of partitional and hierarchical clustering algorithms." Data clustering, Chapman and Hall/CRC, 87-110, https://doi.org/10.1201/9781315373515.

Reimers, N., and Gurevych, I. (2019). "Sentence-bert: Sentence embeddings using siamese bert-networks." arXiv preprint arXiv:1908.10084, https://doi.org/10.48550/arXiv.1908.10084.

Sacks, R., Radosavljevic, M., and Barak, R. (2010). "Requirements for building information modeling based lean production management systems for construction." Automation in construction, 19(5), 641-655, https://doi.org/10.1016/j.autcon.2010.02.010.

Salama, T., Salah, A., and Moselhi, O. (2021). "Integrating critical chain project management with last planner system for linear scheduling of modular construction." Construction innovation, 21(4), 525-554, https://doi.org/10.1108/CI-05-2018-0046.

Salling, S. T., Perez, C. T., and Wandahl, S. (2023). "Perception of Project Management Among Construction Workers: A Survey in Denmark." Proc., Proceedings of the 31st Annual Conference of the International Group for Lean Construction, 882-893, https://doi.org/10.24928/2023/0124.

Seppänen, O., Ballard, G., and Pesonen, S. (2010). "The combination of last planner system and location-based management system." Lean construction journal, 6(1), 43-54, https://www.researchgate.net/publication/228417459.

Sheikhkhoshkar, M., Bril El-Haouzi, H., Aubry, A., and Hamzeh, F. (2023a). "Functionality as a key concept for integrated project planning and scheduling methods." Journal of Construction Engineering and Management, 149(7), 04023053, https://doi.org/10.1061/JCEMD4.COENG-13427.

Sheikhkhoshkar, M., Bril El-Haouzi, H., Aubry, A., Hamzeh, F., and Rahimian, F. (2024a). "From process-based to technology-driven: a study on functionalities as key elements of collaborative planning methods for construction projects." Production Planning & Control, 1-22, https://doi.org/10.1080/09537287.2024.2360581.

Sheikhkhoshkar, M., Bril El Haouzi, H., Aubry, A., and Hamzeh, F. (2024b). "An advanced exploration of functionalities as the underlying principles of construction control metrics." Smart and Sustainable Built Environment, https://doi.org/10.1108/SASBE-12-2023-0379.

Sheikhkhoshkar, M., El-Haouzi, H. B., Aubry, A., Hamzeh, F., and Poshdar, M. (2023b). "Analyzing the lean principles in integrated planning and scheduling methods." Proc., Proc., 31st Annual Conference of the International Group for Lean Construction, IGLC31, https://doi.org/10.24928/2023/0159.

Silveira, B. F., and Costa, D. B. (2023). "Method for automating the processes of generating and using 4d bim models integrated with location-based planning and last planner® system." Construction Innovation, https://doi.org/10.1108/CI-02-2022-0030.

Viana, D. D., Formoso, C. T., and Isatto, E. L. (2017). "Understanding the theory behind the Last Planner System using the Language-Action Perspective: two case studies." Production planning & control, 28(3), 177-189, https://doi.org/10.1080/09537287.2016.1233360.

Wang, K., Reimers, N., and Gurevych, I. (2021). "Tsdae: Using transformer-based sequential denoising auto-encoder for unsupervised sentence embedding learning." arXiv preprint arXiv:2104.06979, https://doi.org/10.48550/arXiv.2104.06979.

Wiendahl, H.-H., Von Cieminski, G., and Wiendahl, H.-P. (2005). "Stumbling blocks of PPC: Towards the holistic configuration of PPC systems." Production Planning & Control, 16(7), 634-651, https://doi.org/10.1080/09537280500249280.

Zeng, N., Ye, X., Liu, Y., and König, M. (2023). "BIM-enabled Kanban system in construction logistics for real-time demand reporting and pull replenishment." Engineering, Construction and Architectural Management, https://doi.org/10.1108/ECAM-01-2022-0036.

Zhao, X., Yeoh, K.-W., and Chua, D. K. H. (2020). "Extracting construction knowledge from project schedules using natural language processing." Proc., The 10th International Conference on Engineering, Project, and Production Management, Springer, 197-211, 9811519099, https://doi.org/10.1007/978-981-15-1910-9_17.

Zhou, S., Ng, S. T., Lee, S. H., Xu, F. J., and Yang, Y. (2019). "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.

Zohrehvandi, S., Khalilzadeh, M., Amiri, M., and Shadrokh, S. (2021). "Project buffer sizing and dynamic buffer consumption algorithm in power generation construction." Engineering, Construction and Architectural Management, 29(2), 716-738, https://doi.org/10.1108/ECAM-08-2020-0605.

Zulu, S. L., Saad, A. M., and Gledson, B. (2023). "Exploring leaders' perceptions of the business case for digitalisation in the construction industry." Buildings, 13(3), 701, https://doi.org/10.3390/buildings13030701.

APPENDIX A

Table A1: Extracted functionalities from the literature.

ID	Functionalities
F1	Achieving an integrated cost/schedule progress monitoring and control
F2	Aligning goals with the owner
F3	Aligning the work plan assignment with the look-ahead
F4	Allowing just-in-time (JIT) purchasing and delivery of material
F5	Allowing the interactions of multiple team professionals and stakeholders in a common environment.
F6	Analyzing constraints more effectively
F7	Analyzing schedule constructability
F8	Applying analytical method
F9	Applying creativity techniques to planning and scheduling
F10	Assisting in reducing waste
F11	Automated planning of concrete joint layouts
F12	Automatic generation of the as-built real-time 4D simulation
F13	Automatically determines the resource quantities
F14	Automating the generation of schedule
F15	Avoiding omissions and sequencing mistakes
F16	Better flow of information
F17	Considering contractual requirement
F18	Considering Supply chain instability
F19	Considering the continuous flow of work
F20	Constructability evaluation
F21	Continuous improvement process
F22	Continuous learning
F23	Controlling of uncertainty
F24	Controlling project progress and performance
F25	Controlling the cost of the project in progress
F26	Creativity, option generation, and innovation
F27	Decentralized work tracking and information communication on construction sites
F28	Decreasing meeting durations
F29	Decreasing workflow variability
F30	Detected more logical errors, more accurately, and faster, with less need for intrateam communication
F31	Detecting and solving spatiotemporal conflicts
F32	Dynamic collision detection and spatial-temporal conflict analysis
F33	Early involvement of key project stakeholders
F34	Easing of access and low training time
F35	Effective supply chain practices
F36	Eliminating the root causes of variability
F37	Empowering automated project progress monitoring
F38	Enabling lean construction adoption and situation awareness
F39	Enabling project performance prediction
F40	Enabling real-time collaborative 4D planning to gain a robust construction plan
F41	Enabling real-time communication with workers
F42	Enabling real-time tracking in construction site
F43	Enabling the coordination of the look-ahead plans
F44	Enabling value management/engineering
F45	Enhancing the performance of construction project management
F46	Enhancing transparency
F47	Evaluating the performance of the look-ahead level
F48	Experiencing and reviewing scheduled sequences on a 1:1 scale
F49	Facilitating communication with subcontractors
F50	Filling gaps related to delay analysis
F51	Forecasting project duration
F52	Fosters the convergence of 4D uses with project documents
F53	Generating real-time interactive project visualization
F54	Higher sense of immersion and interaction
F55	Highlighting the importance of short-term planning at the crew level
F56	Identifying and eliminating wastes
F57	Identifying highly sensitive activities
F58	Identifying possible optimizations
F59	Identifying root causes for deviations.
F60	Identifying time-space conflicts
F61	Implementing pull flow control
F62	Improving continuous learning
F63	Improving alignment of engineering & procurement with construction and commissioning
F64	Improving collaborative sensemaking

Table A1: (continued).

ID	Functionalities
F65	Improving communication and teamwork between the project team
F66	Improving decision-making among geographically dispersed industry practitioners
F67	Improving engineering curriculum
F68	Improving managerial practices
F69	Improving organizational agility
F70	Improving predictability
F71	Improving production tracking, forecasting and control
F72	Improving role description and responsibilities
F73	Improving safety management on-site
F74	Improving the involvement and commitment of all professional groups
F75	Improving the reliability of the planning
F76	Improving the understanding of the project's progress
F77	Improving the usability of the 4D BIM for workflow analyses
F78	Increasing productivity
F79	Increasing safety on construction sites
F80	Increasing teamwork, and communication between all stakeholders
F81	Increasing the work-flow reliability onsite
F82	Increasing transparency
F83	Integrating offsite and onsite planning for modular and offsite construction
F84	Linking the supply chain and construction process
F85	Maintaining a workable backlog
F86	Maintaining continuity of resources
F87	Maintaining continuity of trade work
F88	Maintaining production rate stability
F89	Making decisions under a user-centric approach
F90	Managing constraints removal
F91	Managing corrective actions
F92	Managing uncertainties
F93	Managing work density
F94	Measuring labour productivity
F95	Measuring labour resource reliability
F96	Measuring long-term and short-term plan alignment
F97	Measuring reliability and effectiveness of weekly work plan and look-ahead plan
F98	Measuring the achievability of a target project duration
F99	Measuring the efficiency of resource allocation
F100	Measuring the make-ready process
F101	Measuring the percentage of differentiated tasks in the WWP
F102	Measuring the quality of capacity planning
F103	Measuring the quality of the commitments
F104	Measuring the quality of the construction flow
F105	Measuring the reliability and stability of the workflow
F106	Monitoring worker motion and worker location
F107	More efficient resource management.
F108	More natural and industry-adapted interactions during a collaboration session
F109	Moving participants from passive receivers of a schedule to active contributors to the schedule
F110	Permitting real-time virtual collaboration for stakeholders from different locations
F111	Promoting the reliability of the Commitment Plan
F112	Providing predictive control/monitoring methods to make the best decisions with forwarding simulation
F113	Providing a digital equivalence to face-to-face communication in construction projects
F114	Providing better budget control
F115	Providing enhanced awareness of ongoing work
F116	Providing ergonomic interactions with the session workflow.
F117	Providing managerial information
F118	Providing more valuable handouts to the planning meetings' participants
F119	Pull planning effectiveness
F120	Quantifying the reliability of starting or finishing the task on time
F121	Real-time safety monitoring
F122	Reducing information loss in data exchange
F123	Reducing the change orders during construction
F124	Reducing the impacts of potential causes of delay
F125	Reducing uncertainty and risk
F126	Reducing production cycle time
F127	Relating workflow reliability to productivity
F128	Reshaping an individual’s cognitive determinants to influence collaboration throughout project delivery
F129	Safer learning environment
F130	Scheduling of modular and offsite construction

Table A1: (continued).

ID	Functionalities
F131	Sharing the knowledge and lessons learned
F132	Site layout and environment management
F133	Solving disputes more efficiently
F134	Solving site logistic problems
F135	Supporting co-navigate, co-sort, co-plan, co-simulate and co-talk
F136	Supporting collaboration through social conversations
F137	Supporting human decision-making
F138	Supporting multi-functional review meetings
F139	Supporting the learning and process of off-site construction production planning challenges
F140	Supporting the trust and reliable promises among the team
F141	Systematic identification and removal of constraints
F142	Testing research hypotheses and improving understanding of the schedule
F143	Understanding of the subprocesses more accurately
F144	Understanding the behaviour of the performance indicators
F145	Visualizing alternative construction sequences based on various what-if scenarios
F146	Visualizing the schedules to understand and communicate content to a variety of stakeholders
F147	Visualizing the status of work-in-progress
F148	Workspace planning