Advanced Engineering Informatics 57 (2023) 102079

Contents lists available at ScienceDirect

Advanced Engineering Informatics

journal homepage: www.elsevier.com/locate/aei

Full length article

Lexicon-based content analysis of BIM logs for diverse BIM log mining use cases

Suhyung Jang, Ghang Lee *, Sanghyun Shin, Hyunsung Roh

Building Informatics Group, Department of Architecture and Architectural Engineering, Yonsei University, Republic of Korea

ARTICLE INFO

Keywords:
building information modeling (BIM)
BIM log mining
BIM log
Lexicon-based content analysis
Supportability

ABSTRACT

This study conducted a lexicon-based content analysis of building information modeling (BIM) logs from four major BIM authoring tools and four custom-developed BIM loggers to understand whether the BIM logs satisfy the information requirements for various BIM log mining use cases, as well as to assess their potential for future development and research. First, through a critical review of previous studies, 19 different ways of using BIM logs were identified, including authoring and collaborative pattern discovery, authoring process modeling, collaboration pattern analysis, command predictions, and team optimization; however, most of the uses concerned process discovery. The analysis also revealed that BIM log mining has mainly been used for the design phase, with a few examples of being used for the construction phase. For BIM log mining, various techniques ranging from simple frequency analyses via social network analyses to advanced pattern discovery were deployed. In terms of BIM log sources, native BIM logs from Revit were dominantly used almost in all studies, aside from a few studies that used custom-developed BIM logs. The content analysis of BIM logs showed that the contents of native BIM logs provided by major BIM authoring tools varied, but commonly lacked model-element-specific information; this limitation prevents in-depth analyses of BIM processes. Overall, the current disproportionate focus on process discovery in the design phase of BIM log mining suggests that the application of BIM log mining is still in its early stages and holds significant potential for other project phases if adding model-element-specific information is incorporated.

1. Introduction

A building information modeling (BIM) log is a sequential record of events automatically collected during the use of BIM software. With the increasing adoption of BIM, BIM logs have become an abundant and valuable source of information for improving modeling practices and processes. Although many BIM log mining studies have been conducted, previous research has not fully investigated the readiness of BIM logs for diverse mining use cases. Furthermore, previous studies have predominantly concentrated on the usage of logs from one particular BIM software application, namely Revit, while less attention has been given to other widely used BIM software tools in the architecture, engineering, construction, and operations (AECO) industry. Thus, this study aimed to examine the contents of BIM logs from various BIM authoring tools, as well as custom-developed BIM loggers; in this way, we assessed the extent to which current BIM logs can support the information required for various BIM log mining use cases and identify future research and development directions.

To achieve this goal, this study first conducted a comprehensive review of the current state of BIM log mining, focusing on the information requirements of objectives and contents of BIM logs. The review was conducted through a reproducible screening process of previous BIM log mining studies and BIM logs from various sources, including major BIM authoring tools such as Allplan, Archicad, Revit, and Vectorworks [1]. Through the literature review, the log mining objectives, techniques, and information requirements for each analysis objective were derived. Then, a lexicon-based content analysis was conducted to examine the information items included in the native logs collected from the major BIM authoring tools. Based on this analysis, the authors of this study further evaluated the extent to which the BIM logs fulfill the information

Abbreviations: AD, authoring characteristic discovery; AM, authoring process modeling; ANN, artificial neural network; API, application programming interface; CC, construction conformance checking; CD, collaborative characteristic discovery; CM, construction process modeling; CP, authoring command prediction; DC, design collaboration conformance checking; OT, optimal team configuration; PP, construction process prediction.

* Corresponding author.
E-mail addresses: rgb000@yonsei.ac.kr (S. Jang), glee@yonsei.ac.kr (G. Lee), spicysh03126@yonsei.ac.kr (S. Shin), hs-roh0928@yonsei.ac.kr (H. Roh).

https://doi.org/10.1016/j.aei.2023.102079
Received 21 March 2023; Received in revised form 30 May 2023; Accepted 26 June 2023
Available online 1 July 2023

requirements of the cases. The evaluation results were comparatively analyzed to assess the current state of BIM log mining and provide suggestions for future directions.

The remainder of the article is organized as follows: Section 2 briefly reviews previous BIM log mining studies and discusses research gaps. Section 3 presents the terminology and research methodology employed in this study. Section 4 provides a literature review on the screened literature, analyzing BIM log mining use cases and custom BIM loggers proposed in the previous studies. Section 5 conducts a lexicon-based content analysis to examine the presence of information items in the BIM logs recorded by major BIM authoring tools. Section 6 evaluates the supportability of BIM logs for the various use cases. Section 7 discusses the current state of this field and proposes future directions for readers' consideration. Finally, Section 8 concludes this paper.

2. Previous BIM log mining studies and research gaps

Pioneering research in BIM log mining was conducted by Yarmohammadi et al. [2] in 2016. This foundational work analyzed the Revit journal file, a type of BIM log, yielding insights into productive design patterns. This paved the way for a series of studies that focused on analyzing BIM logs to understand and improve BIM processes. Several analytical techniques have since been applied to BIM logs, including sequential pattern mining [2,3], attribute-based clustering [4–6], graph-based network analysis [7–10], and sequence prediction algorithms [11–13]. These techniques aim to mine knowledge and optimize processes using the ever-growing repositories of BIM logs. As the field of BIM log mining continues to advance, there is a need for deeper exploration into the critical factors that impact outcomes, in line with insights from process mining. In particular:

Information requirements of use cases: Process mining hinges on the quality and relevance of the event log contents [14]. Bose et al. [15] identified four main categories of issues impacting event log quality based on an examination of over a hundred process mining projects: missing data, incorrect data, imprecise data, and irrelevant data. Suriadi et al. [16] further refined these issues into 11 event log imperfection patterns, along with corresponding remedies to rectify them. Rozinat and van der Aalst [17] proposed an incremental approach to assess the fitness and suitability of event logs. Eck et al. [18] emphasized the importance of aligning research questions with project goals and ensuring that the data content can sufficiently address these questions. However, although such advancements in process mining have improved the understanding of information requirements and improved the content and quality of event logs, this aspect has been underexplored in BIM log mining. Our study emphasizes the alignment of information contained within BIM logs with the requirements of defined research questions and use cases.

Contents of BIM logs: Previous studies have made progress in appending required information items for desired analyses by developing custom BIM loggers [9,10,19,20,37]. However, these studies did not include detailed content analyses of BIM logs. Most previous researchers have acknowledged the limitations of BIM logs in accurately representing the BIM process and the associated challenges in analyzing their unstructured data format [3,9,10,19]. Furthermore, previous studies have primarily focused on Revit BIM logs [6], without considering the potential variations in log content from different BIM authoring tools. This research expands on this concept by evaluating the extent to which BIM logs from various major BIM authoring tools can fulfill the diverse information requirements of BIM log mining use cases.

This study aims to deepen the understanding of BIM log mining by analyzing use cases and data requirements, analyzing BIM logs generated by major BIM authoring tools, and examining the extent to which the contents of BIM logs fulfill the information requirements of BIM log mining use cases. To address the current research gaps, this study formulates three research questions:

(1) "What are the use cases of BIM log mining and their information requirements?".
(2) "What are the contents of the BIM logs?".
(3) "To what extent do the contents of BIM logs fulfill the information requirements, and thereby offer support, for each identified use case?".

By addressing these questions, this study aims to augment the current understanding of BIM log mining and outline potential future directions.

3. Research method

This study consists of three main components, as depicted in Fig. 1: (1) a literature review, (2) a lexicon-based content analysis, which is an approach grounded in text mining techniques designed to comprehend and parse unstructured text data [21], and (3) a supportability evaluation. Each component corresponds to one of the three research questions we posed. In the next subsection, relevant terminology related to BIM log mining is reviewed to establish a common understanding of the key concepts. The following subsections detail the methodology we used for the literature review of past BIM log mining use cases, the lexicon-based content analysis of BIM logs, and the evaluation of the supportability of BIM logs for different use cases.

3.1. Definitions of "BIM log" and "BIM log mining"

In previous studies on BIM log mining, there has been an inconsistent understanding of the concept due to the varying terminology used to describe data sources and analysis methods, as shown in Table 1.

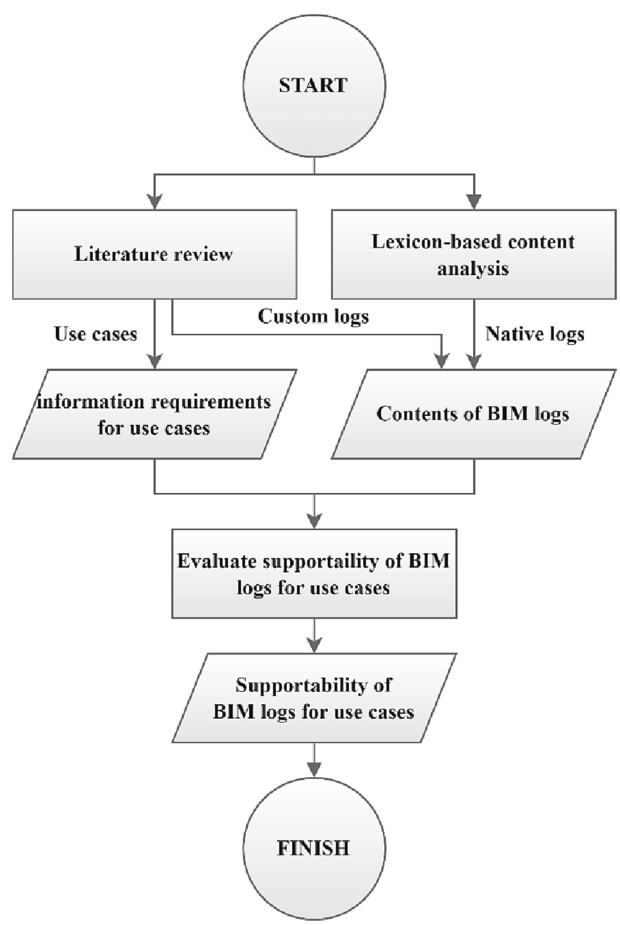

Fig. 1. Research flow.

Table 1
Terms used for analysis methods and data sources in previous studies.

Analysis methods	Data sources	References
Data analytics	Detailed model development data	Yarmohammadi and Castro-Lacouture [19]
Data mining	Design log file	Yarmohammadi et al. [3]
Log data mining	BIM design log	Pan and Zhang [22]
Log data mining	BIM log file	Forcael et al. [23]
Log data mining	BIM event log	Pan and Zhang [12]
Log mining	Event log	Zhang and Ashuri [7]
Event log mining	BIM event log	Pan et al. [4], Pan et al. [5], Pan and Zhang [25]
Process mining	Event log	Kouhestani and Nik-Bakht [9]
Process mining	BIM event log	Pan and Zhang [10], Pan and Zhang [26], Pan and Zhang [24]
Process mining	3D modeling event log	Gao et al. [20], Gao et al. [13]
Process mining	BIM log file	Zhang et al. [27]
BIM log mining	BIM log	Ishizawa [6]

Although previous studies employed different terms for data sources, such as "BIM log file," "detailed model development data," or "BIM event log," the term "BIM log" is used in this study to collectively refer to these data sources. As for analysis methods, some early studies treated BIM log mining as a type of data analytics technique [3,19] encompassing various data analysis approaches, including log mining. However, other studies specifically categorized BIM log mining as a type of log mining [7,12,22,23]. The majority of studies [4,5,9,10,13,20,24] considered BIM log mining as an extension of process mining (i.e., event log mining). According to Wil van der Aalst and his colleagues, process mining involves the use of "techniques, tools, and methods to discover, monitor, and improve real processes (i.e., not assumed processes) by extracting knowledge from event logs commonly available in today's information systems" [14]. This study regards BIM log mining as a refined application of process mining, uniquely adapted for BIM logs.

The following definitions for "BIM log," "BIM log mining," and other related terms are used in this study:

● A *BIM* log is a chronological record of events that occurred during the use of BIM software, such as design authoring, including additions, changes, and deletions.

● A native (BIM) log refers to a BIM log recorded by an off-the-shelf logger provided by a BIM software vendor.

● A custom (BIM) log refers to a BIM log recorded by a custom-developed logger.

● *BIM* log *mining* refers to techniques to discover, monitor, and improve processes in the AECO industry by extracting knowledge from BIM logs.

3.2. Paper screening and selection

The scope of the literature review includes the following: (1) use case analyses, (2) information requirements for use cases, and (3) contents of custom logs. The authors of this study collected scientific publications from July 2016 to the date on which the search was conducted (March 20, 2023) using representative academic databases including Scopus, Web of Science (WoS), American Society of Civil Engineers (ASCE) Library, and Wiley Online Library. The commencement period of July 2016 was selected because it coincides with a notable milestone in the field, that is, the first BIM log mining study by Yarmohammadi et al. [2]. The literature screening procedure of this study followed a three-step paper screening procedure, depicted in Fig. 2. In the first step, search queries and conditions were determined to automatically identify academic papers related to BIM. The existing literature was retrieved by using the keywords "building information model*". The wildcard character "*" was used to capture relevant variations of a word, such as "building information model," "building information modelling," and "building information modeling". The abbreviation "BIM" was omitted to avoid articles using different meanings for this abbreviation, such as Budget Impact Model and Bcl-2 Interacting Mediator. To limit search results to relevant sources, this study used the query "BIM log" OR "event log" OR "log data" OR "log file" instead of using "log". The word "log" has multiple meanings in scientific papers, such as a piece of lumber, a tool for measuring a ship's motion, or an abbreviation of

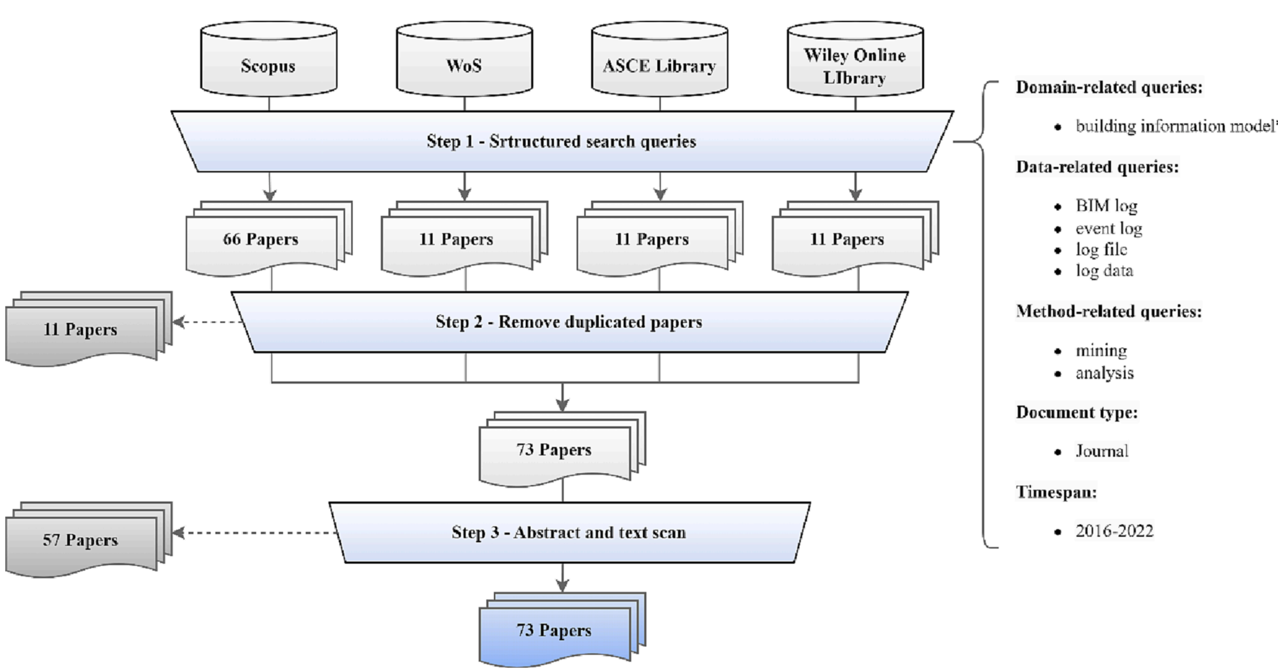

Fig. 2. Paper screening process.

"logarithm." In particular, papers related to the AECO industry contain a large volume of articles related to timber structures that use the first aforementioned meaning of "log". This study restricted the document type to peer-reviewed journal articles published in English, as they are both reputable and reliable [28]. A total of 68, 11, 10, and 4 papers were obtained from Scopus, WOS, ASCE Library, and Wiley Online Library, respectively. In the second step, duplicated papers were removed. This study compared titles, author names, and digital object identifiers to identify the duplicates. A total of 12 papers were removed, leaving 81 papers. In the third step, the authors manually reviewed the remaining papers to screen for irrelevant articles. The authors read the abstracts and main text to identify articles that did not fit the objective of the literature review, which was to identify unique BIM log mining use cases. Six articles from different research areas were excluded, and 58 articles that did not use BIM logs as a source of analysis were also excluded. As a result, the third step excluded 64 irrelevant papers and a total of 17 papers remained for review, as listed in Table 2.

Use cases were analyzed based on three aspects: analysis type, objective, and techniques. The authors initially identified unique techniques for specific objectives (i.e., use cases) and categorized them according to three process mining analysis types [14]: process discovery, conformance checking, and process enhancement. Process discovery aims to uncover insights about a process through analysis of event logs, while conformance checking seeks to identify inconsistencies between a process and its corresponding event log. Process enhancement aims to improve or extend a process based on the insights gained from analyzing the event log. The distributions of these analysis types are a significant indicator for the maturity of process mining applications. Van der Aalst regarded conformance checking and process enhancement as more mature process mining applications compared to process discovery [30]; the review provided by Garcia et al. [31] indicated that process discovery is dominant in early stages of process mining adoption, with the proportions of conformance checking and process enhancement growing as the application field broadens. Consequently, this paper investigated BIM log mining research trends by comparing the distributions of analysis types. Additionally, research objectives, which represent specific phases and domain tasks in BIM log mining applications, were examined. Lastly, techniques applied in BIM log mining were briefly reviewed to provide insights into potential future research developments. Information requirements for each use case were examined to establish information requirements. To standardize the semantics of information items with designations that varied across studies, the authors established a standard set of eight information items commonly used in previous studies. These items include *timestamp*, a digital record of the occurrence of a particular event; *user ID*, an identifier distinguishing the user; *session ID*, an identifier for the period from BIM software initiation to closure; *project name*, a user-assigned project designation; *modeling command*, a predefined action executable within the BIM software; *element ID*, an identifier for a BIM element within the project; *element attribute value change*, an updated value of BIM element property derived from a modeling command; and *element category*, a predefined classification of BIM elements within the software (e.g., walls, columns, doors, etc.). Upon analyzing the information requirements of the aforementioned information items in selected previous BIM log mining use cases, we established the information requirements specific to each use case.

Table 2
Screened papers.

Date	Authors	Title
SEP 2017	Yarmohammadi et al.	Mining implicit 3D modeling patterns from unstructured temporal BIM log text data [3]
JAN 2018	Zhang et al.	BIM Log Mining: Measuring Design Productivity [27]
MAY 2018	Yarmohammadi and Castro-Lacouture	Automated performance measurement for 3D building modeling decisions [19]
JUL 2018	Zhang and Ashuri	BIM log mining: Discovering social networks [7]
JAN 2020	Pan and Zhang	BIM log mining: Exploring design productivity characteristics [22]
MAR 2020	Forcael et al.	Behavior and Performance of BIM Users in a Collaborative Work Environment [23]
APR 2020	Kouhestani and Nik-Bakht	IFC-based process mining for design authoring [9]
APR 2020	Pan and Zhang	BIM log mining: Learning and predicting design commands [12]
JUL 2020	Pan et al.	Clustering of designers based on building information modeling event logs [4]
DEC 2020	Pan et al.	Mining event logs for knowledge discovery based on adaptive efficient fuzzy Kohonen clustering network [5]
JAN 2021	Pan and Zhang	A BIM-data mining integrated digital twin framework for advanced project management [10]
JUL 2021	Pan and Zhang	Automated process discovery from event logs in BIM construction projects [26]
OCT 2021	Gao et al.	A data structure for studying 3D modeling design behavior based on event logs [20]
JUL 2022	Ishizawa	Keystone players in collaborative building information modeling—form of contribution in Japanese large-scale projects [6]
OCT 2022	Gao et al.	Command prediction based on early 3D modeling design logs by deep neural networks [13]
OCT 2022	Pan and Zhang	Modeling and analyzing dynamic social networks for behavioral pattern discovery in collaborative design [25]
JUN 2023	Gao et al.	Impact of 3D modeling behavior patterns on the creativity of sustainable building design through process mining [29]

Additionally, the information items included in four custom logs developed in previous studies were analyzed [9,10,13,19]. The scope of analysis included identifying the presence of these eight information items in the custom logs. Only information items explicitly cited within the articles were taken into account for the analysis.

3.3. Lexicon-based content analysis method

This subsection introduces the lexicon-based content analysis employed in this study, also referred to as a keyword in context analysis [32], to analyze the contents of the BIM logs written by the major BIM authoring tools: Allplan, Archicad, Revit, and Vectorworks. A lexicon is a comprehensive set of predefined keywords that are used as a common basis for comparing and analyzing different terms used in text data. The lexicon-based content analysis assesses the presence and frequency of the keywords in text data [32–34], and is therefore an effective method for analyzing text data with an unknown structure. The lexicon-based content analysis was executed in four steps: data collection, data preprocessing, lexicon development, and content analysis.

A simulated BIM authoring process was designed for data collection. The process comprised eight BIM authoring activities, as listed in

Table 3
Activities of the simulated BIM authoring process.

Activity	Description	Contents of interest
A_1	Start a session and create a new project.	[User ID, Project name]
A_2	Add a default wall.	[BIM command, Element ID, Element category]
A_3	Modify wall height to 2850 mm.	[BIM command, Element ID, Element attribute change]
A_4	Add a default door on the wall.	[BIM command, Element ID, Element category]
A_5	Change the library of the door.	[BIM command, Element ID, Element attribute change]
A_6	Delete the door.	[BIM command, Element ID]
A_7	Undo the deletion.	[BIM command]
A_8	Save the project and close the session.	[User ID, Project name]

Table 3, commonly available in the BIM authoring tools. These activities included initiation or termination of a project session (A_1 and A_8); adding (A_2 and A_4), modifying (A_3 and A_5), and deleting (A_6) building elements; and redoing a command (A_7). The activities encompass all BIM authoring activities covered in the previous BIM log mining studies. The data collected in this this step included a BIM model, a native log, and a manual record, as shown in Fig. 3. The BIM model was produced through the BIM authoring process, whereas the native log was automatically generated by the BIM authoring tools. The manual record, which recorded the initiation and completion times of each authoring activity, was recorded by the authors during the simulated BIM authoring process.

In the data preprocessing step, events in the native logs were isolated by activity. Owing to the nature of the lexicon-based approach, it was essential to prevent events from different activities from mixing within a single text file, as this method relies on target text data and a lexicon for analysis. To isolate the events related to a specific BIM authoring activity, the authors filtered the logs using the initiation and completion times recorded in the manual record, as shown in Fig. 4. Owing to the lack of *timestamp* in Archicad's native log, it was isolated manually instead of relying on the records. The Archicad's native log was duplicated into a separate folder after each execution of an activity. After completion of the authoring process, the consequent versions were compared to isolate the events appended during an activity.

In the lexicon development step, the authors developed a lexicon for each BIM authoring activity for all BIM authoring tools, as shown in Fig. 5. The authors first investigated the designations for information items in each BIM authoring tool. In this process, *timestamp* and *session ID* were excluded, as they were identified based on a checklist during the following content analysis. The designations for *modeling command* and *project name* were not separately analyzed because they are generally accepted concepts among all.

BIM authoring tools, however; the values of each were investigated. The authors further examined the value of each information item from the BIM models created using the corresponding BIM authoring tool. The scope of investigation for each activity was the same as the contents of

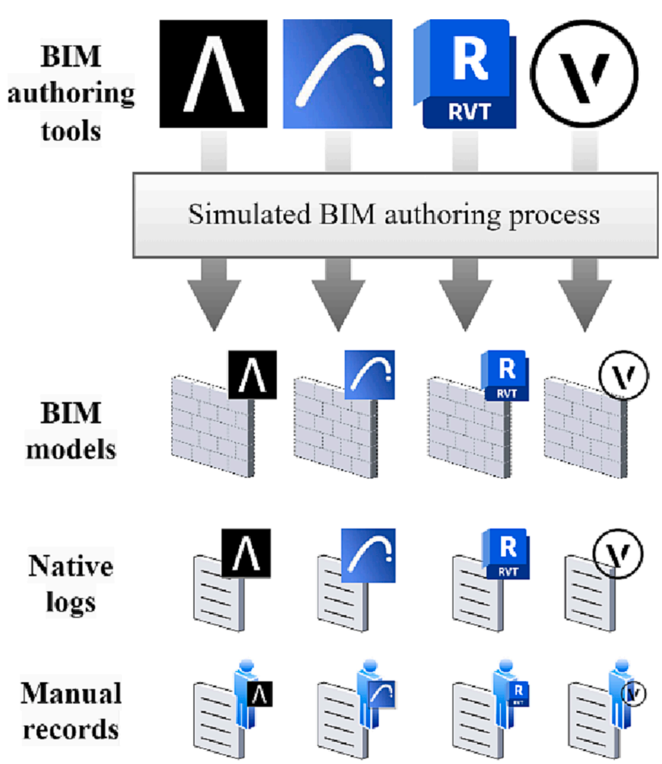

Fig. 3. Step 1: Data collection.

interest listed in Table 3. For example, activities A_1 and A_8 initiate or terminate a project session run by the user; therefore, the relevant information items for these activities were *project name* and *user ID*. Activities A_2 and A_4 involve modeling commands for adding therefore relevant information items include *modeling command*, *element ID*, and *element category,* whereas the deleting activity, A_6, includes *modeling command* and *element ID*, and the redo activity, A_7, only includes *modeling command.* Likewise, activities A_3 and A_7 modify the attributes of elements; therefore, relevant information items are *modeling command, element ID*, and *attribute change.* Because the BIM authoring tools use different designations for each information item, the authors manually investigated their equivalent values by examining each tool and the values of keywords within the BIM models to establish the lexicon. In some cases, there were multiple concepts that complied with the definition of an information item, in which case all corresponding values of the keywords were included as a list.

In the content analysis step depicted in Fig. 6, BIM logs were analyzed using a combination of a manual checklist and a keyword search algorithm. The checklist was employed to identify the presence of timestamps and session IDs, which were manually marked by the authors. The authors visually examined the logs to identify the presence of *timestamp*. Additionally, it was assessed whether BIM logs were saved separately for each execution of the BIM authoring tools to determine the presence of *session ID*. For native logs not saved by session, it was determined whether there was any distinguishable information between the sessions. The keyword search algorithm assesses each isolated log for the presence of predefined keywords (i.e., lexicon) of other information items. The algorithm iterates through the logs and their corresponding lexicons to identify the presence of the keywords in each log written during the activities. The inclusion of information items was determined using a combination of logical disjunction and logical conjunction. *User ID* and *project name* were evaluated using logical disjunction, as they are not event-dependent information items. *Modeling command*, *element ID*, *element attribute value change*, and *element category*, however, were evaluated using logical conjunction, as they are event-dependent information items.

Through the four aforementioned steps, this approach allowed the authors to analyze BIM logs with unknown contents and formats. Although this study analyzed the native logs of four representative BIM authoring tools, the proposed approach is generally applicable for other BIM logs whose contents and formats are not known.

3.4. Supportability of BIM logs for BIM log mining use cases

This study developed a concept of supportability to evaluate the extent to which the contents of the BIM logs fulfill the information requirements of use cases. This use of supportability and other relevant terms are defined as follows:

● A *use case* (u) is a unique application of BIM log mining techniques to achieve a specific objective, such as design productivity analysis or monitoring project progress.

● An *information item* (i) is a data item that describes a property of a BIM log event, such as its identifier (ID), timestamp, name, or other data.

● An *information requirement* (R_u) is a set of *information items* (i_1, i_2, …, i_n) that are required by a specific *use case* (u) for a BIM log.

● *Contents* (C_l) are a collection of *information items* (i_1, i_2, …, i_n) recorded in a BIM log (l).

● *Supportability S*(l, u) indicates the extent to which the *contents* (C_l) of a *BIM* log (l) fulfill the *information requirements* (R_u) of a specific *use case* (u). Supportability can be formally defined as in Equation 1 below:

$$S(l,u) = \begin{cases} 1(Supported), if \dfrac{|C_l \cap R_u|}{|R_u|} = 1 \\ 0(Notsupported), otherwise \end{cases}$$

Equation 1. Supportability of a BIM log for a use case.

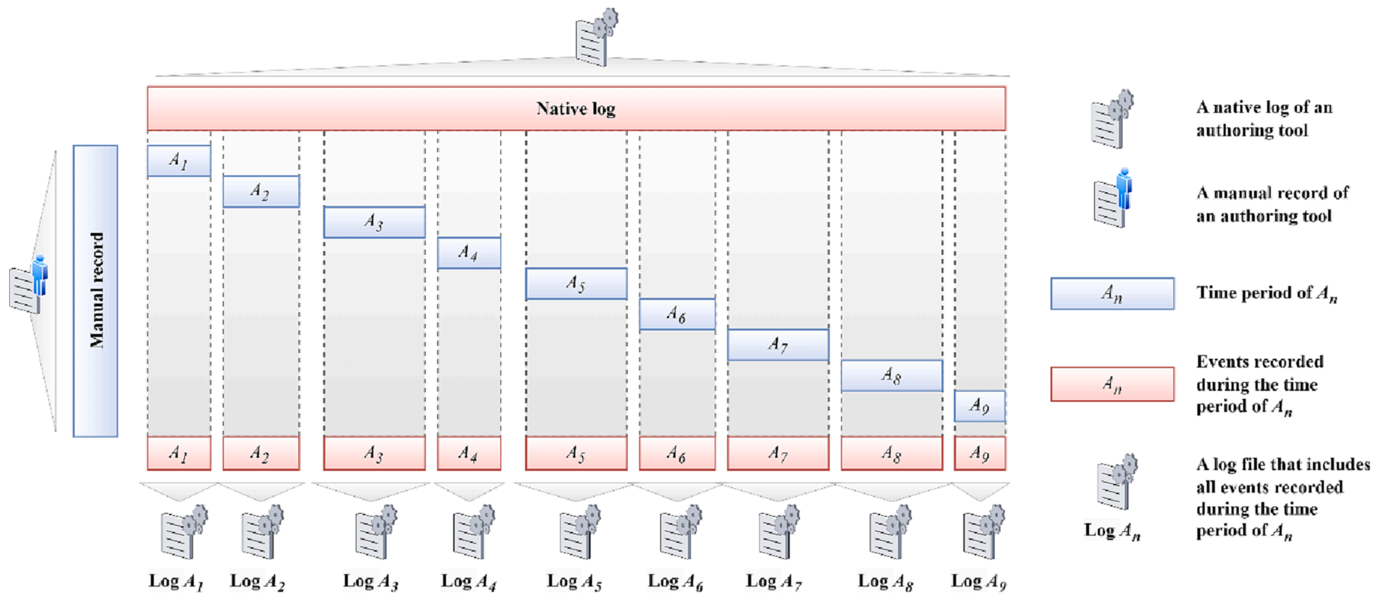

Fig. 4. Step 2: Data preprocessing.

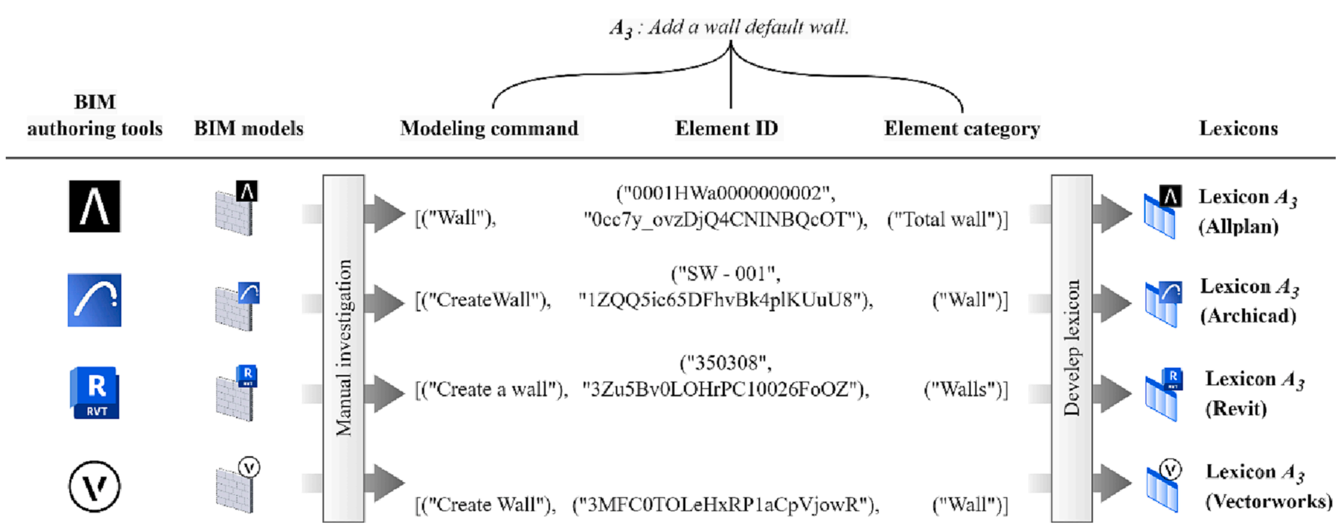

Fig. 5. Step 3: Lexicon development.

The supportability of eight BIM logs for BIM log mining use cases was evaluated based on the information requirements for the use cases and the contents of the BIM logs based on the results of the literature review and the lexicon-based content analysis. To facilitate a comprehensive evaluation, the authors grouped the BIM logs into two categories based on their data sources (native and custom logs) and grouped the BIM log mining use cases according to their objectives and analysis types. Average supportability values were calculated for each group and a discussion is provided to offer an overarching interpretation of the results of this study, highlighting the current state of BIM log mining and identifying areas for future improvement.

4. BIM log mining use cases, techniques, and information requirements

4.1. BIM log mining use cases

The authors first reviewed the full text of the selected articles to analyze three aspects of the BIM log mining use cases: analysis types, objectives, and techniques. Yarmohammadi et al. [3] and Zhang et al. [27] conducted early studies that analyzed the performance of designers based on the most frequent command patterns extracted from BIM logs. Yarmohammadi and Castro-Lacouture were the first to develop a custom log using the Revit application programming interface (API) to optimize team configuration based on the productivity of each designer on different authoring tasks [19]. Zhang and Ashuri analyzed the features of a design firm and the role of individual designers based on the network of designer collaboration [7]. Pan and Zhang investigated time-dependent productivity by analyzing the performance characteristics of the designers, clustering them based on variations in productivity across specific hours of the day and days of the week [22]. Forcael et al. assessed the behavior and performance of BIM users in a collaborative environment, using BIM logs to identify patterns, preferred commands, experience levels, training needs, and strategies for improving team performance. [23]. Kouhestani and Nik-Bakht proposed a method using industry foundation class (IFC)-based custom logs for the process modeling of BIM authoring, enabling conformance checks, network analysis, and bottleneck analysis [9]. Another study by Pan and Zhang developed a framework using long short-term memory neural networks (LSTM NNs) to predict design commands based on BIM logs, aiming to

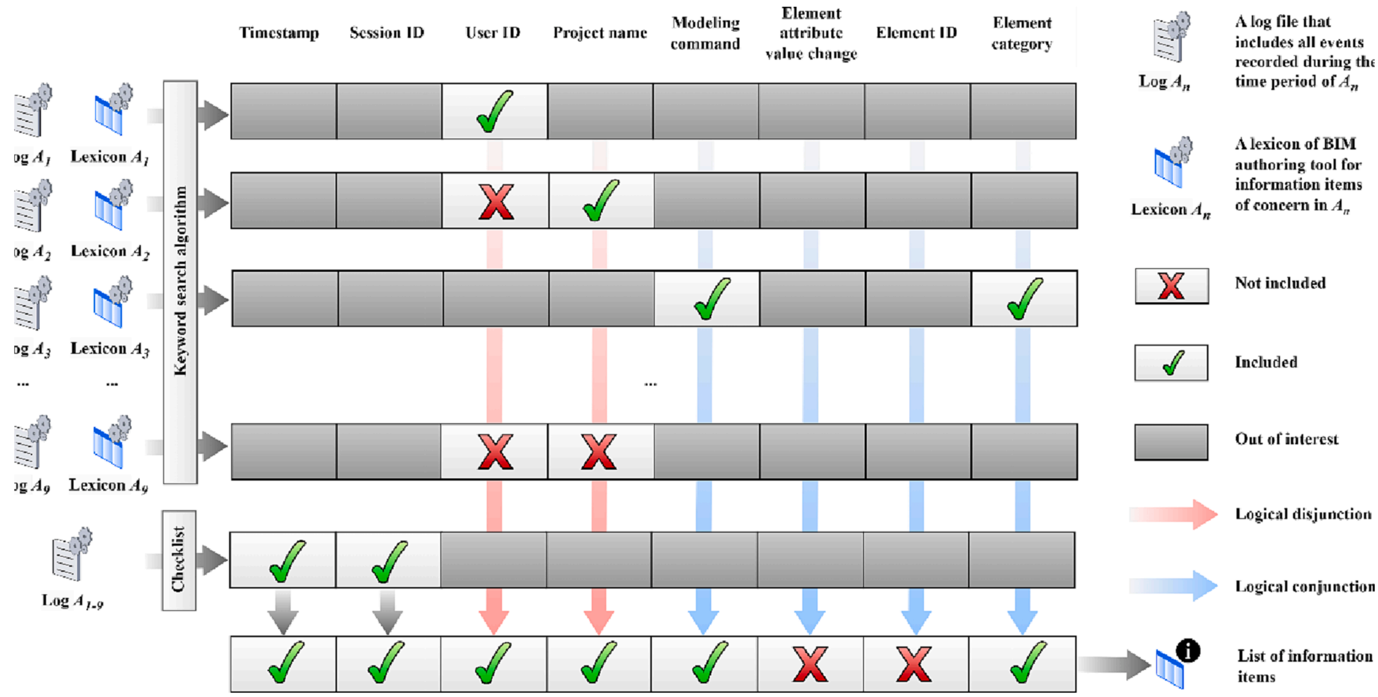

Fig. 6. Step 4: Content analysis.

Table 4
BIM log mining use cases: analysis types, objectives, and techniques.

Analysis types	Objectives	Acronyms	Use cases	Techniques	References
Process Discovery	Authoring characteristic discovery (AD)	AD1	Extraction and evaluation of frequent modeling patterns and their efficiency.	Sequence mining algorithm	[3,27]
		AD2	Clustering of designers based on productivity characteristics.	Clustering algorithm	[5,22]
		AD3	Analysis of command distribution by discipline.	Statistical comparison	[23]
		AD4	Clustering of designers based on their command contribution ratios.	Clustering algorithm	[6]
		AD5	Extraction and evaluation of creative design behavior patterns.	Neural network-based prediction model	[29]
	Collaborative characteristic discovery (CD)	CD1	Analysis of the networks of design firms and roles of individual designers.	Network analysis	[7]
		CD2	Analysis of time-varying networks of design firms and roles of individual designers.	Time-varying network analysis	[25]
		CD3	Clustering of designers based on collaborative patterns.	Network analysis and clustering	[4]
		CD4	Analysis of design collaboration between disciplines in a design project.	Network analysis	[9]
		CD5	Analysis of design collaboration between participants in a construction project.	Network analysis	[26]
	Authoring process modeling (AM)	AM1	Identification of bottlenecks in the BIM authoring process.	Inductive miner algorithm	[9]
		AM2	Normalization of BIM authoring processes by element categories.	Graph simplification algorithm	[20]
	Construction process modeling (CM)	CM1	Identification of bottlenecks in the BIM construction process.	Inductive miner algorithm	[10,26]
Conformance Checking	Design collaboration conformance (DC)	DC1	Evaluation of BIM design process conformance.	Comparison between as-happened and as-planned design process using inductive miner algorithm.	[9]
	Construction conformance (CC)	CC1	Evaluation of BIM construction process conformance.	Comparison between as-happened and as-planned construction process using inductive miner algorithm.	[26]
Process Enhancement	Authoring command prediction (CP)	CP1	Prediction of modeling commands.	Sequence prediction deep learning models	[12]
		CP2	Prediction of modeling commands for specific object categories.	Sequence prediction deep learning models	[13]
	Construction process prediction (PP)	PP1	Prediction of construction progress.	Statistical sequence prediction model	[29]
	Optimal team configuration (OT)	OT1	Determination of the optimal team configuration.	Earliest due date (EDD) sequencing method	[19]

improve modeling efficiency and automate the design process [12]. Pan et al. employed unsupervised machine learning techniques to analyze collaboration [4] and productivity characteristics [5] in BIM projects, enhancing design processes and efficiency while fostering teamwork and collaboration for project managers. Pan and Zhang proposed a BIM-data mining integrated digital twin framework for advanced project management [10], which was further developed to a method for automated process discovery from event logs in BIM construction projects [26]. Gao et al. presented a data structure for studying 3D modeling design behavior based on BIM logs [20] and employed a deep neural network-based approach for modeling command prediction [13]. Ishizawa analyzed the frequencies of contributory commands to cluster the key players within a firm [6]. Pan and Zhang proposed using dynamic social network analysis to objectively evaluate the performance of designers and predict their influence, aiding data-driven decision-making for optimized collaboration within a design firm [25]. Lastly, Gao et al. explored a neural network-based approach to identify and evaluate creative design behavior patterns in sustainable building design based on BIM logs [29].

A total of 19 use cases of BIM log mining were identified from previous studies, as listed in Table 4. The 19 use cases were first grouped into three analysis types corresponding to commonly used top-level categories of process mining [14], namely process discovery, conformance checking, and process enhancement. The three analysis types were further categorized into eight objectives, after which the 19 use cases were coded based on these objectives. Full details are presented in Table 4.

Fig. 7 shows the distribution of the use cases by analysis types and objectives. At the analysis type level, the process discovery category includes over 70 percent of the use cases. At the objective level, 5 out of 19 use cases focused on AD and 5 focused on CD, followed by 2 out of 19 use cases focusing on AM and 2 focusing on CP. Among the use cases, AD1, AD2, and CM1 were applied in two articles, while the other use cases were identified as corresponding to one article each. As previously mentioned, process discovery was the most dominant analysis type, with few use cases were conducted for conformance checking (DC1 and CC1) and process enhancement (CP1, CP2, PP1, and OT1). The distribution of the analysis types identified in this study aligns with that in the early application period as analyzed by Garcia et al. [31], in that use cases of process discovery were dominant, whereas use cases of process enhancement and conformance checking steadily became more common as the technology matured. The BIM log mining application was dominant for the design phase, including 15 of the 17 articles; two articles focused on the construction phase, proposing four use cases, namely CD5, CM1, CC1, and PP1 [10,26]. The techniques applied in BIM log mining varied by analysis type and objective.

Among the analysis types, process discovery involved the widest variety of applied techniques. For AD, pattern mining algorithms, clustering algorithms, and statistical comparisons were applied to understand the efficiency of the authoring process based on frequent command patterns (AD1), productive or contributory characteristics of the designers (AD2 and AD4), and analysis of command patterns by discipline (AD3). Among the process discovery use cases, AD5 employed an artificial neural network (ANN) approach, which was a unique use of neural networks within the category of process discovery. CD was conducted using various network analysis techniques to understand the collaborative characteristics of designers and design firms. The specific techniques applied in the studies include network density analysis (CD1), centrality analysis (CD1, CD2, and CD3), clustering of networks (CD3), and traces of work handovers between participants within a project (CD4 and CD5). AM and CM were conducted through process abstraction. Specifically, AM1 and CM1 were conducted using the inductive miner from ProM, a popular process mining tool, and AM2 was conducted using a graph simplification algorithm. Conformance checking was conducted in both DC1 and CC1 to compare the as-planned process with the as-happened process. DC1 employed visual comparison and an inductive miner to assess alignment with BIM guidelines, whereas CC1 applied the inductive miner to BIM logs for the as-happened process to analyze deviation between the two processes. Process enhancement use cases employed both prediction models and domain-specific approaches. In CP1, a deep learning-based approach was utilized with an LSTM NN to develop a prediction model and compare its performance against other machine learning models such as K-nearest neighbor, random forest, and support vector machines. CP2, on the other hand, compared the performance of several neural network models (bi-directional LSTM, LSTM-NN, sequence to sequence (Seq2-Seq), Seq2Seq with attention mechanism, gated recurrent units, and transformer). PP1 applied a statistical sequence prediction model, the multivariate autoregressive integrated moving average model, to predict the number of finished construction tasks, thus enabling advanced project management and decision-making to optimize construction operations. OT1 was conducted using the earliest due date (EDD) sequencing rule in combination with the critical path method (CPM), which are both traditional concepts for project management in the AECO industry.

The literature review of articles on BIM log mining reveals a range of

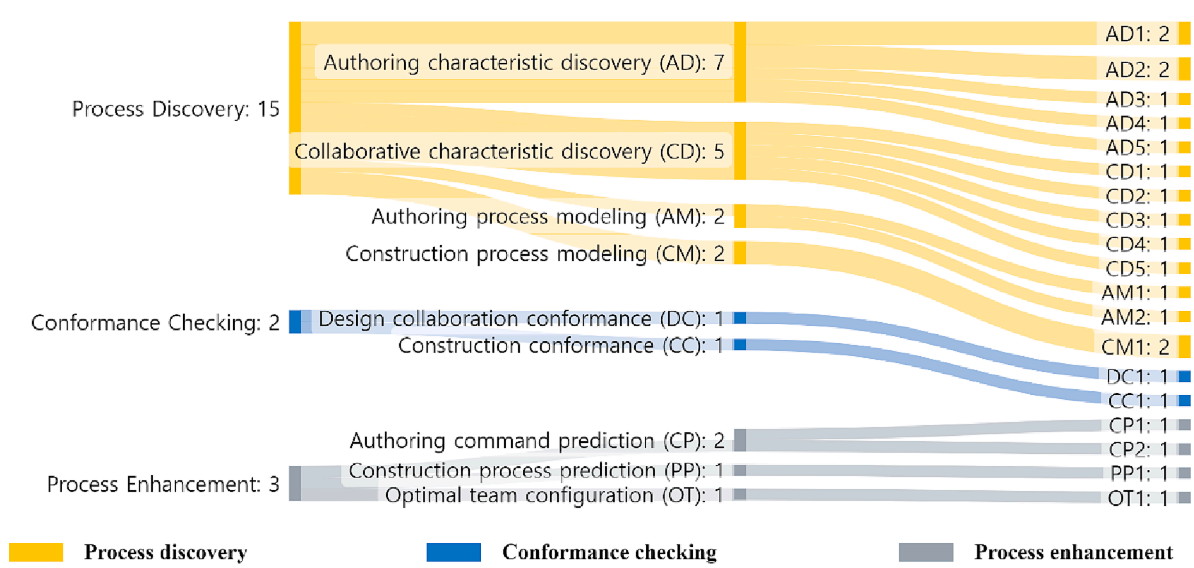

Fig. 7. Distribution of BIM log mining studies by research objective.

use cases for BIM data analysis, with process discovery being the predominant type of analysis. The techniques used for analysis vary depending on the type and objective of the analysis; for example, process discovery employs techniques like pattern mining, clustering, and network analysis, whereas process enhancement uses prediction models and domain-specific approaches, such as EDD and CPM. This review indicates an expanding use of advanced deep learning techniques like ANN and LSTM NN, which are broadening research objectives and opening up new potential applications. In addition, BIM log mining has shown potential for integration with traditional process analysis methods in the AECO industry. Although BIM log mining has been predominantly employed for analysis of the design phase, it is gradually becoming more commonly utilized in the construction phase. Thus, BIM log mining is still in its early stages in terms of the number of studies and areas of application, but it holds significant potential for future growth in improving operations and decision-making throughout the life cycle of the AECO project.

4.2. Information requirements

Table 5 outlines the standard set of eight information items required to conduct the 19 research use cases based on previous studies on BIM log mining. *Timestamp* was the most commonly used in 17 out of 19 use cases, followed by *user ID*, and *modeling command*, *element category*, and *element ID*, used in 16, 14, 12, and 11 use cases, respectively. These results align with the information requirements of general process mining, which requires a timestamp, case ID, and activity as essential information items [14].

In terms of the timestamp, CD1 did not utilize *timestamp* in its network construction. Instead, the network of the design firm was constructed based on the project name without considering a chronological order of collaborative activities. In terms of the case ID, previous studies have used various approaches, with some using *user ID* as the sole case ID (7 out of 19), whereas others used *element ID* as the sole case ID (2 out of 19). Additionally, more than half of the studies (10 out of 19) used both *user ID* and *element ID*. In the case of CP1, however, the *modeling command* itself was used as a case ID for authoring command prediction, without utilizing a separate case ID. In terms of activity, certain studies utilized *element attribute value change* instead of *modeling command* as activity. Specifically, CD5, CM1, CC1, and PP1, which aimed to apply BIM log mining in construction management, utilized *element attribute value change*. This was because the state of construction per element was recorded as the element attribute and its values, and the BIM logs utilized in the studies focusing on construction management were recorded by comparing sequential snapshots of monitored construction BIM models.

This analysis reveals that BIM log mining relies on the prerequisites of process mining, namely the timestamp, activity, and case ID [14]. However, a significant difference exists between process mining and BIM log mining. Process mining aims to generalize the overall process to identify discrepancies or activities that require improvement. In contrast, the BIM process is unique to a specific project [35]. Therefore, BIM log mining can focus on the performance of participants who are repeatedly involved in the projects or the development of building elements of a certain category throughout the project life cycle [20]. This focus can also be inferred from the frequent use of *user ID* and *element category* in the use cases. This fundamental difference between BIM log mining compared to traditional process mining can be identified from the analysis of the information requirements.

4.3. Contents of custom logs

The information items included in the custom logs introduced in the previous studies were also analyzed using the literature review. The result of the analysis is summarized as Table 6. Yarmohammadi and Castro-Lacouture were the first to propose a custom BIM logger developed using Revit API [19]. In their study, they analyzed the average time taken to model each *element category*. This was done by evaluating the *timestamp* interval between execution of *modeling commands* on *element IDs* throughout the authoring process. The study utilized *timestamp*, *element ID*, and *element category* to discriminate modeling efficiency differences between individual modelers, then jointly utilized EDD and CPM to attain the optimal team configuration given the number of building elements and their categories. Even though it was not utilized in the proposed method, their custom logger includes *element attribute value changes* such as “element type”, “element family”, and “bounding box”, as well as *user ID* and *project name*. Their log files were saved separately by session and were therefore regarded to include *session ID*.

Kouhestani and Nik-Bakht developed an IFC logger by comparing a snapshot of BIM models recorded during the BIM process [9]. The framework was divided into two steps; in the first step, a version of the BIM model of the ongoing process was saved whenever the model size exceeded the threshold. In the second step, sequential snapshots were compared to detect changes in BIM elements. The information items recorded from the changes include *timestamp*, resource ID (i.e., actors or *user ID*), *modeling command*, *element attribute value change* (such as attachment, area, volume, structural usage, and other), *element ID*, and IFC element category (i.e., *element category*). Although it was not mentioned in the article, the snapshot-based algorithm introduced in their article does not consider the session of the modeling process, but

Table 5
Information items required by each objective.

Use cases	Timestamp	Session ID	User ID	Project name	Modeling command	Element attribute value change	Element ID	Element category
AD1	✓		✓		✓			
AD2	✓		✓		✓			
AD3	✓		✓		✓			
AD4	✓		✓		✓			
AD5	✓		✓				✓	✓
CD1			✓	✓	✓			
CD2	✓	✓	✓		✓			
CD3	✓	✓	✓		✓			
CD4	✓		✓		✓		✓	✓
CD5	✓		✓			✓	✓	✓
AM1	✓		✓		✓		✓	✓
AM2	✓		✓		✓		✓	✓
CM1	✓		✓			✓	✓	✓
DC1	✓		✓		✓		✓	✓
CC1	✓		✓			✓	✓	✓
CP1	✓				✓			✓
CP2	✓				✓		✓	✓
PP1	✓		✓			✓	✓	✓
OT1	✓		✓		✓		✓	✓

Table 6
Contents of custom loggers.

Custom logger	Timestamp	Session ID	User ID	Project name	Modeling command	Element attribute value change	Element ID	Element category
Yarmohammadi and Castro-Lacouture	✓	✓	✓	✓	✓	✓	✓	✓
Kouhestani and Nik-Bakht	✓	✓	✓		✓	✓	✓	✓
Pan and Zhang	✓	✓	✓		✓	✓	✓	✓
Gao et al.	✓	✓	✓		✓		✓	✓

designates logging directories by project; therefore, *session ID* was considered to be missing and *project ID* was considered to be included.

Pan and Zhang further enhanced the IFC logger by incorporating construction task-specific attributes [10,26] using SYNCHRO, which is BIM software for virtual construction planning and operations. Element attributes including "IfcClass", "TaskID", "TaskName", "TaskStart", "TaskFinish", and "Participant", related to the classification of construction tasks, their durations, and the workers involved, were embedded in IFC files and compared according to the procedures proposed by Kouhestani and Nik-Bakht. Unlike the IFC logger for authoring process analysis, which redefined *modeling command* based on the attribute change, the IFC logger for construction management focused on retrieving *element attribute value changes*. Other information items including *timestamp*, trace ID (i.e., *user ID*), *element ID*, and *element category* remained the same. In addition, the inclusion of *session ID* and *project ID* were considered to be the same as in the previously developed IFC logger.

Gao et al. developed a modeling logger using Rhino API [13,20,29] to record *timestamp*, user name (i.e., *user ID*), *modeling command*, and globally unique identifier (GUID) (i.e., *element ID*). The *element category* was categorized as glaze wall, vertical panel, and horizontal mullion based on layer names [13,20] or the object names defined in Rhino [29]. As their logger was manually initiated and terminated during modeling sessions, their custom log was regarded to include *session ID*.

The analysis of the contents of the custom loggers reveals that the custom logs contain various information items, including *timestamp*, *user ID*, *modeling command,* and *element category*, which were already known to exist in the Revit native log. However, custom loggers capture additional information items, such as *element ID* and *element attribute value change*. Although the custom BIM loggers were developed using different software or development interfaces, such as the Revit C# API, Dynamo, and Rhino API, common efforts were made to record chronological changes in building elements. When comparing the content of custom logs with the findings from the use case requirement analysis, custom logs distinguished themselves from known native logs (i.e., Revit journal file) by including model-element-specific information items. However, these custom logs could face inherent challenges due to their nature. They might not be publicly accessible and might necessitate manual updates, potentially limiting the efficacy of BIM log mining. As such, the authors of this study explored the contents of other native logs, as described in the following section.

5. Lexicon-based content analysis

This section details the results of the lexicon-based content analysis, which involved data collection, data preparation, lexicon development, and content analysis. The versions of the BIM authoring tools utilized in this study were as follows: Allplan 2023, Archicad 26, Revit 2023, and Vectorworks 2023 SP1 in the English language and on Windows operating systems to ensure reproducibility and up-to-date native logs. The remainder of this section presents the step-by-step results of the lexicon-based content analysis, focusing on the differences between the BIM authoring tools in terms of their BIM logs and contents.

5.1. Data collection

The authors of this research collected four unique datasets from simulated BIM authoring processes that were carried out on Allplan, Archicad, Revit, and Vectorworks. Each dataset contained a BIM model, a native log, and a manual record. The file sizes of the BIM models and native logs, coupled with the number of lines written in the native logs, are depicted in Fig. 8. These data highlight the considerable variations in the events recorded in the native logs across different BIM authoring tools, even when the same BIM authoring process is employed.

To illustrate this point, the BIM models created by Revit and Vectorworks showed comparable sizes, with the Revit model at 5,556 kilobytes and the Vectorworks model at 5,236 kilobytes. However, the Revit native log size was approximately three times larger than that of Vectorworks, with a size of 310,806 bytes versus 99,625 bytes. Additionally, the volume of information recorded for each event showed a noticeable disparity among the BIM authoring tools. Although the native logs produced by Archicad and Revit had similar sizes, 288,074 bytes and 310,806 bytes, there was a significant difference in the number of lines each log contained. Archicad's log had 8,780 lines, whereas Revit's contained just 3,232 lines. These results underscore the inherent variability in event recording and information capture among different BIM authoring tools, with respect to both file size and information volume, even when the same authoring process is followed.

5.2. Data preparation

The native logs collected from the simulated BIM authoring process were isolated into partial logs by activity using initiation and completion times recorded in manual records. Fig. 9. presents the file size proportions of the isolated logs by activity. It was observed that the proportions of events recorded during preparation (A_1) and closure (A_8) of the software were dominant in the native logs of Archicad and Revit, which had larger native log sizes compared to the other two BIM authoring tools. The authors identified three most frequently used word token in the Archicad and Revit native logs, excluding exclamation marks. In the Archicad native log, "Library", "Files\\Graphisoft\\Archicad", and "26\\Archicad" were included 724, 541, and 395 times respectively, whereas in the Revit native log, "API_SUCCESS", "Used", and "application" were included 655, 440, and 433 times, respectively. The frequency of words in the native logs can provide insight into activities recorded by the BIM authoring tools. For example, frequent recordings of "Library" and specific file paths in the Archicad native logs suggest that they were written to monitor data loading for stable system operation. Similarly, frequent recordings of "API_SUCCESS" and verification of the application in the Revit native logs align with their original purpose of operational management [36]. The eight partial logs isolated from the native BIM logs of each authoring tool were analyzed in the content analysis step, along with the results of the following lexicon development.

5.3. Lexicon development

The designations for information items in each BIM authoring tools are listed in Table 7. For *user ID*, the authors identified two different

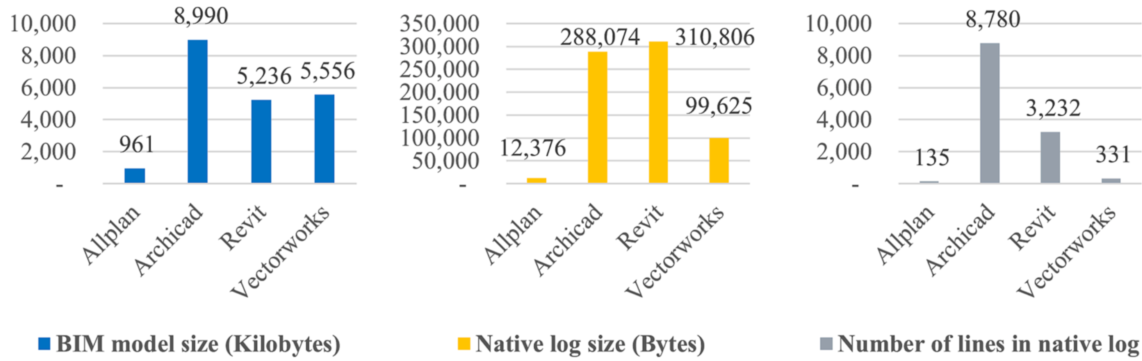

Fig. 8. Sizes of collected datasets.

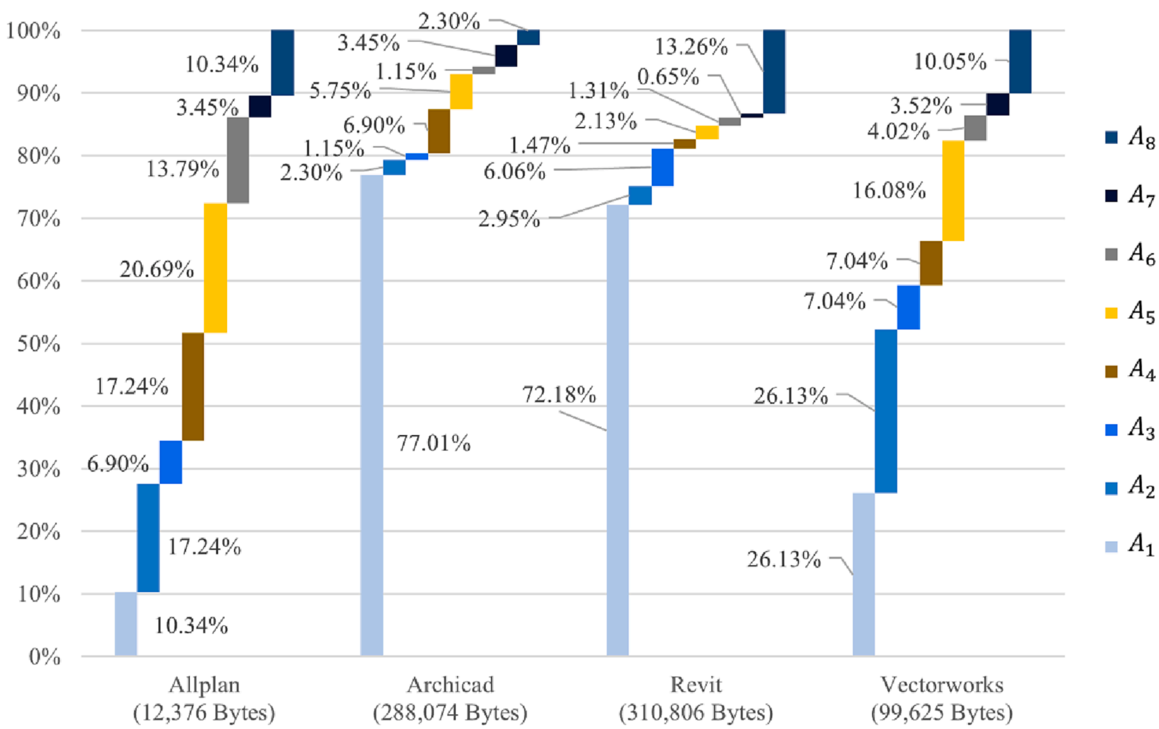

Fig. 9. Log sizes by activity.

Table 7
Designations for information items in each BIM authoring tool.

BIM authoring tool	User ID	Project name	Element attribute value change (A_3)	Element attribute value change (A_5)	Element ID	Element category
Allplan	License, User	File name, Document	Height	Library	Allright_Comp_ID, Ifc ID	Object_name
Archicad	Serial, registered name	File name	Height	Library	ID, Archicad IFC,	Classifi-cation
Revit	username	File name	Unconnected Height	Type	Element ID, IfcGUID	Category
Vectorworks	Serial Number, Account name	Fine name	Height	Library	GlobalId	Class

concepts that comply with the definition: license serial and account name. License serial protects software intellectual property and prevents unauthorized distribution while facilitating troubleshooting and customer support. Allplan, Archicad, and Vectorworks manage licensing by serials referred to as “License,” “Serial,” and “Serial Number,” respectively. Archicad, Revit, and Vectorworks have adopted account-based licensing systems for the versions used for this study, referring to the account name as “registered name,” “username,” and “Account name,” respectively. Revit provides network licensing in which account-based licensing is solely used without providing license serials. For *element ID*, the authors identified two different concepts that comply with the definition: IfcGloballyUniqueId and software internal ID. IfcGloballyUniqueId holds an encoded string identifier that is used to uniquely identify an IFC object and is supported in all four BIM authoring tools. It is referred to as “Ifc ID,” “Archicad IFC,” “IfcGUID,” and “GlobalId” in Allplan, Archicad, Revit, and Vectorworks, respectively. The software internal ID is an encoded identifier that is either automatically or manually given to elements by the modelers. Allplan

Table 8
Lexicons of BIM authoring tools by activities.

Activity	Contents of interest	Allplan	Archicad	Revit	Vectorworks
A_1	[(User ID),(Project name)]	[(“XXX-XX-XX-XXXXX”, “rgb000@yonsei.ac.kr”), (“Allplan_model.ndw”)]	[(“XXXXXXXXXXXXXXXXXXXX”,“rgb000@yonsei.ac.kr”),(“Archicad_model.pln”)]	[(“rgb000@yonsei.ac.kr”),(“Revit_model.rvt”)]	[(“XXXXXX-XXXXXX-XXXXXX-XXXXXX”,“rgb000@yonsei.ac.kr”), (“Vectorworks_model.vwx”)]
A_2	[(BIM command), (Element ID), (Element category)]	[(“Wall”), (“0001HWa0000000002”, “0cc7y_ovzDjQ4CNINBQcOT”), (“Total wall”)]	[(“CreateWall”), (“SW – 001”,“1ZQQ5ic65DFhvBk4plKUuU8”), (“Wall”)]	[(“Create a wall”), (“350308″,“3Zu5Bv0LOHrPC10026FoOZ“), (“Walls“)]	[(“Create Wall”), (“3MFC0TOLeHxRP1aCpVjowR”), (“Wall”)]
A_3	[(BIM command), (Element ID), (Element attribute change)]	[(“Change”), (“0001HWa0000000002”, “0cc7y_ovzDjQ4CNINBQcOT““), (“2850”)]	[(“Change Wall”), (“SW – 001”,“1ZQQ5ic65DFhvBk4plKUuU8”), (“2850”)]	[(“Modify | Walls”), (“350308”,“3Zu5Bv0LOHrPC10026FoOZ”), (“2850”)]	[(“Shape Pane Edit”), (“3MFC0TOLeHxRP1aCpVjowR”), (“2850”)]
A_4	[(BIM command), (Element ID), (Element category)]	[(“Door”), (“0001Sym0000000135”, “1uUk$Uoxf2YgWfM9jiw_6b”), (“Door, smart symbol + SmartPart”)]	[(“InsertDoor”), (“DOO – 001”,“0znj5hb1b25gB8WvqeNEIS”), (“Door”)]	[(“Create a door”), (“350458”,“3W2Tu2KKr66wtPU$rPGsDK”), (“Doors”)]	[(“Create Object”), (“2v5SSflOKHxRPKaCpVjowR”], (“Door”]]
A_5	[(BIM command), (Element ID), (Element attribute change)]	[(“Change”), (“0001Sym0000000135”, “1uUk$Uoxf2YgWfM9jiw_6b”), (”1| white 1 side”)]	[(“Change Door”], (“DOO – 001”,“0znj5hb1b25gB8WvqeNEIS”), (“Entrance Double Door 26”)]	[(“Modify | Doors”), (“350458”,“3W2Tu2KKr66wtPU$rPGsDK”), (“M_Single-Flush”)]	[(“Shape Pane Edit”], (“2v5SSflOKHxRPKaCpVjowR”), (“Door Hinged 2P”)]
A_6	[(BIM command), (Element ID)]	[(“Delete”), (“0001Sym0000000135”, “1uUk$Uoxf2YgWfM9jiw_6b”)]	[(“Delete”), (“DOO – 001”, “0znj5hb1b25gB8WvqeNEIS”)]	[(“Delete”),(“350458”,”3W2Tu2KKr66wtPU $rPGsDK”)]	[(“Delete”), (“2v5SSflOKHxRPKaCpVjowR”)]
A_7	[(BIM command)]	[(“Undo”)]	[(“Undo”)]	[(“Undo”)]	[(“Undo”)]
A_8	[(User ID),(Project name)]	[(“XXX-XX-XX-XXXXX”, “rgb000@yonsei.ac.kr”), (“Allplan_model.ndw”)]	[(“XXXXXXXXXXXXXXXXXXXX”,“rgb000@yonsei.ac.kr”)],(“Archicad_model.pln”)]	[(“rgb000@yonsei.ac.kr”),(“Revit_model.rvt”)]	[(“XXXXXX-XXXXXX-XXXXXX-XXXXXX”,“rgb000@yonsei.ac.kr”), (“Vectorworks_model.vwx”)]

and Revit have “Allright_Comp_ID” and “element ID” as the internal ID, which are automatically given by the software and not allowed to be changed by the users. Archicad and Vectorworks have “ID” and “classification identification” for the internal ID; these can be modified by the author. However, the “ID” of Archicad is given automatically, whereas “classification identification” has to be given manually. Because assigning the *element ID* was not included in the simulated authoring process, “classification identification” was exempted from the scope of lexicon development.

The element attributes that need to be identified in A_3 and A_5 for *element attribute value change* are “wall height” and “element library”. An “element library” is a collection of pre-made 3D models, objects, and components that can be repeatedly used in BIM modeling process. All BIM authoring tools except for Revit utilized the same designations for the concepts, namely “Height” and “Library”, whereas Revit designated it as “Unconnected Height” and “Type”. The *element category* is designated differently among all BIM authoring tools, namely “Object_name”, “Classification”, “Category”, and “Class”, respectively, for Allplan, Archicad, Revit, and Vectorworks.

Table 8 presents the lexicons with corresponding values as keywords for the information items investigated in the BIM models. Lexicons were developed using a list structure with outer levels distinguished by “[]” and inner levels distinguished by “()”. Inner-level lists were used for information items with multiple values, such as.

user ID and element ID. License serial characters were replaced with “X” and account names were replaced with “rgb000@yonsei.ac.kr”. Project names were given by the authors using the rule “<Tool name>_model. < Extension of BIM model from the tool>”. The results of the lexicon development reveal that the standard set of information items is valid for all BIM authoring tools, although the designations and the formats of the values were different among the BIM authoring tools.

5.4. *Content analysis*

Table 9 presents the results of the content analysis of Allplan. The native log of Allplan only contained *timestamp* and *session ID* among the eight information items. The *timestamp* of the Allplan native log was recorded in the “YYYY-MM-DD hh:mm:ss [.fraction]” format. Although the native log of Allplan was written as a single file that was stacked over multiple sessions, the Allplan native log contained the statements “***** Start of General Allplan Log” and “End of General Allplan Log *****” at the beginning and ending of each session. Therefore, the *session ID* was considered to be included. However, the native logs of A_3, A_5, A_6, and A_7 in Allplan included designations for the modify, cancel, and redo modeling commands, but not for element creation, which was a commonly used type of information item in the previous studies. Hence, *modeling command* in the Allplan native log was deemed missing.

Table 10 presents the results of the content analysis of Archicad. The native log of Archicad contained all information items except for the *timestamp*, *element ID*, and *element attribute value change*. Among the BIM authoring tools, the Archicad native log was the only one that did not include the *timestamp*. However, because the Archicad native log was written separately for each session, the native logs of Archicad were considered to include the *session ID*. The *user ID* of Archicad was identified in A_1 based on inclusion of the account name, but the *project name* was only identified in A_8. Because Archicad creates the new document without a given name during A_1, it is given in A_8 during saving and closing.

Table 11 presents the results of the content analysis of Revit. The Revit native log contained most of the information items, except for the *element ID*. The *timestamp* of Revit was recorded in the “DD-Mmm-YY hh:mm:ss [.fraction]” format, with the month recorded as three-character abbreviations (e.g., Jan, Feb, Mar, etc.). The native logs were saved separately for each session and named by *session ID*, which was counted as included. Unlike in the previous study, it was found that the Revit native log contained information about *element attribute value change* and

Table 9
Content analysis results of Allplan.

Activity	Timestamp	Session ID	User ID	Project name	Modeling command	Element attribute value change	Element ID	Element category
A_1								
A_2								
A_3					✓			
A_4								
A_5					✓			
A_6					✓			
A_7					✓			
A_8								
Checklist	✓	✓						
Total	✓	✓	–	–	–	–	–	–

Table 10
Content analysis results of Archicad.

Activity	Timestamp	Session ID	User ID	Project name	Modeling command	Element attribute value change	Element ID	Element category
A_1			✓					
A_2					✓			✓
A_3					✓			
A_4					✓			✓
A_5					✓			
A_6					✓			
A_7					✓			
A_8				✓				
Checklist		✓						
Total	–	✓	✓	✓	✓	–	–	✓

Table 11
Content analysis results of Revit.

Activity	Timestamp	Session ID	User ID	Project name	Modeling command	Element attribute value change	Element ID	Element category
A_1			✓					
A_2					✓			✓
A_3					✓	✓	✓	
A_4					✓			✓
A_5					✓	✓	✓	
A_6					✓			
A_7					✓			
A_8				✓				
Checklist	✓	✓						
Total	✓	✓	✓	✓	✓	✓	✓	✓

element ID in the modifying activities (A_3 and A_5), making it possible to track building element modifications throughout the process; this was not conducted in the previous study.

Table 12 presents the results of the content analysis of Vectorworks. The Vectorworks native log also contained most of the information items, except for the *element ID* and *element attribute value change*. The *timestamp* of the Vectorworks native log was recorded in the "MM/DD/YYYY hh:mm:ss [.fraction]" format. Vectorworks' native log included the *session ID* within the event logs, and it was also considered to include the *session ID*. The *user ID* of Vectorworks was identified both in A_1 and A_8 as inclusion of the license serial, whereas the project name was only identified in A_8, as in Archicad and Revit.

Based on the results, the content of the native logs varied among the different BIM authoring tools. The native logs of Allplan only contained *timestamp* and *session ID*, whereas Archicad did not include the *timestamp*, *element ID*, and *element attribute value change*. Conversely, Revit and Vectorworks contained most of the information items, except for the *element ID* and *element attribute value change* in Vectorworks. Furthermore, the inclusion of *element attribute value change* and partial inclusion of *element ID* were identified in addition to other known information items for Revit. These results show that the employed lexicon-based content analysis was valid in investigating information items from

Table 12
Content analysis results of Vectorworks.

Activity	Timestamp	Session ID	User ID	Project name	Modeling command	Element attribute value change	Element ID	Element category
A_1			✓					
A_2					✓			✓
A_3					✓			
A_4					✓			✓
A_5					✓			
A_6					✓			
A_7					✓			
A_8				✓				
Checklist	✓	✓						
Total	✓	✓	✓	✓	✓	–	–	✓

BIM logs with unknown contents and formats. One commonality identified from the content analysis was the lack of *element ID* and *element attribute value change*, which were identified as significant attributes for understanding the BIM process. The following section will evaluate the supportability of the BIM logs for the BIM log mining use cases, considering both the information requirements for use cases and the contents of the BIM logs.

6. Supportability evaluation

The supportability of the eight BIM logs, comprising four native logs (Allplan, Archicad, Revit, and Vectorworks) and four custom logs (Yarmohammadi and Castro-Lacouture, Kouhestani and Nik-Bakht, Pan and Zhang, and Gao et al.), were evaluated on the 19 objectives analyzed in this study based on the contents of the BIM logs and the information requirements for the objectives. The results of the evaluation are presented in Tables 13, 14, and 15. The authors of this study comparatively analyzed the supportability according to groups of data sources (native and custom logs) and use case categories (analysis types and objectives). Additionally, the analysis of the current state and future directions of BIM log mining is presented in this section.

At the individual BIM log level, the custom logs developed by Yarmohammadi and Castro-Lacouture had the highest supportability (1.00), followed by custom logs developed by Kouhestani and Nik-Bakht. and Pan and Zhang, with an average supportability of 0.95; CD1 requires the project name, which was not supported using their custom logs. The custom log developed by Gao et al. had the lowest average supportability (0.74) among the custom logs, but still showed higher average supportability than any of the native logs. Revit and Vectorworks had the highest average supportability among the native logs (0.42), followed by Archicad (0.05) and Allplan (0.00). Although Archicad contained several information items, it could only support one use case because the *timestamp* was required in most of the use cases. The Allplan log was not applicable to any of the BIM log mining use cases. The analysis revealed that the average supportability of custom logs was 0.91, which was approximately four times higher than the average supportability of native logs (0.22).

At the use case level, use cases AD1 to AD4, CD2, CD3, and CP1 had the highest average supportability (0.75); these can be supported using native logs of Revit and Vectorworks and all custom logs. AD1 to AD4, CD2, and CD3 were supported in six BIM logs, including the native logs of Revit and Vectorworks, as well as all four custom logs. The average supportability values of CD1, CD4, AM1, AM2, DC1, CP2, and OT1 followed (0.5). Among the use cases, all use cases except for CD1 were supported only in the custom logs, whereas CD was supported in the native logs of Archicad, Revit, and Vectorworks. CD5, CM1, CC1, and OT1 showed the lowest supportability (0.38), only being able to be supported using the custom logs developed by Yarmohammadi and Castro-Lacouture, Kouhestani and Nik-Bakht, and Pan and Zhang. At the objective level, AD showed the highest average supportability (0.70), followed by CP (0.63), CD (0.58), AM, DC, and OT (0.5), and CM, CC, and PP (0.38). At the analysis type level, process discovery showed the highest average supportability (0.60), followed by process enhancement (0.53) and conformance checking (0.44).

The results can be summarized as follows. First, missing prerequisites of process mining, such as timestamp, case ID, or activity, can result in low supportability. For example, the Allplan native log did not support any use cases because there were no information items that could be utilized as the case ID or activity in this type of log. Likewise, Archicad could only support CD1, which was the only use case not requiring the *timestamp*. Although Archicad's native log contains *user ID*, which can serve as the case ID, and *modeling command*, which can function as the activity, it lacks a timestamp, a crucial prerequisite for process mining. If Archicad's native log included a timestamp, it could potentially support as many use cases as Revit and Vectorworks, such as AD1 to AD4, CD1 to CD3, and CP1. Second, the inclusion of model-element-specific information items, such as element ID and element attribute value changes, significantly affects BIM log supportability for mature BIM log mining applications. Custom logs generally outperformed native logs in the context of the evaluation, with native logs supporting only 8 out of 19 use cases. Notably, native logs primarily supported process discovery use cases and offered limited support for process enhancement and conformance checking, which are considered as mature applications of process mining. In addition, there were certain use cases that none of the native logs could support, including CD4, CM1, CC1, and PP1, which involved the application of BIM log mining for construction management. This discrepancy resulted from the absence of model-element-specific information items in the native logs. This finding highlights that significant potential for future advancements in BIM log mining lies in understanding the differences between process mining and BIM log mining and incorporating model-element-specific information items in the BIM logs.

Table 13
Supportability of native logs for use cases.

Analysis types	Objectives	Use cases	Native logs			
			Allplan	Archicad	Revit	Vectorworks
Process discovery	**AD**	AD1	0	0	1	1
		AD2	0	0	1	1
		AD3	0	0	1	1
		AD4	0	0	1	1
		AD5	0	0	0	0
	CD	CD1	0	1	1	1
		CD2	0	0	1	1
		CD3	0	0	1	1
		CD4	0	0	0	0
		CD5	0	0	0	0
	AM	AM1	0	0	0	0
		AM2	0	0	0	0
	CM	CM1	0	0	0	0
Conformance checking	**DC**	DC1	0	0	0	0
	CC	CC1	0	0	0	0
Process enhancement	**CP**	CP1	0	0	1	1
		CP2	0	0	0	0
	PP	PP1	0	0	0	0
	OT	OT1	0	0	0	0
Average by BIM logs			0	0.05	**0.42**	**0.42**
Average by data source			**0.22**			

Table 14
Supportability of native logs for use cases.

Analysis types	Objectives	Use cases	Custom logs			
			Yarmohammadi and Castro-Lacouture	Kouhestani and Nik-Bakht	Pan and Zhang	Gao et al.
Process discovery	**AD**	AD1	1	1	1	1
		AD2	1	1	1	1
		AD3	1	1	1	1
		AD4	1	1	1	1
		AD5	1	1	1	1
	CD	CD1	1	0	0	0
		CD2	1	1	1	1
		CD3	1	1	1	1
		CD4	1	1	1	1
		CD5	1	1	1	0
	AM	AM1	1	1	1	1
		AM2	1	1	1	1
	CM	CM1	1	1	1	0
Conformance checking	**DC**	DC1	1	1	1	1
	CC	CC1	1	1	1	0
Process enhancement	**CP**	CP1	1	1	1	1
		CP2	1	1	1	1
	PP	PP1	1	1	1	0
	OT	OT1	1	1	1	1
Average by BIM logs			**1**	0.95	0.95	0.74
Average by data source			**0.91**			

Table 15
Average supportability of BIM logs for use cases.

Analysis types	Objectives	Use cases	Average by use cases	Average by objectives	Average by analysis types
Process discovery	**AD**	AD1	**0.75**	**0.7**	**0.6**
		AD2	**0.75**		
		AD3	**0.75**		
		AD4	**0.75**		
		AD5	0.5		
	CD	CD1	0.5	0.58	
		CD2	**0.75**		
		CD3	**0.75**		
		CD4	0.5		
		CD5	0.38		
	AM	AM1	0.5	0.5	
		AM2	0.5		
	CM	CM1	0.38	0.38	
Conformance checking	**DC**	DC1	0.5	0.5	0.44
	CC	CC1	0.38	0.38	
Process enhancement	**CP**	CP1	**0.75**	0.63	0.53
		CP2	0.5		
	PP	PP1	0.38	0.38	
	OT	OT1	0.5	0.5	

7. Discussion—Current state and future directions

This study aimed to address three research questions concerning the current state of BIM log mining: (1) "What are the use cases of BIM log mining and their information requirements?", (2) "What are the contents of the BIM logs?", and (3) "To what extent do the contents of BIM logs fulfill the information requirements, and thereby offer support, for each identified use case?". The main findings of this study regarding these three research questions can be summarized as follows:

Use cases of BIM log mining and their information requirements: The analysis of BIM log mining studies uncovered diverse use cases for BIM data analysis. Process discovery and process enhancement surface as the leading types of analysis, each employing techniques specific to their objective. The review also hints that the use of advanced deep learning techniques is expanding, suggesting broader research aims and novel potential applications. In terms of information requirements, BIM log mining relies on the prerequisites of process mining, namely the timestamp, activity, and case ID.

However, a clear difference separates BIM log mining from traditional process mining. Unlike process mining, which seeks to generalize the process, BIM log mining is unique to the specifics of each project. It focuses on the performance of participants who are frequently engaged in projects or tracks the development of certain building elements throughout the life cycle of a project. This specific focus is echoed in the frequent use of *user ID* and *element category* in these use cases. Therefore, although BIM log mining borrows some foundational requirements from process mining, it introduces additional unique requirements that better align with its project-specific focus.

Contents of BIM logs: The analysis of custom loggers revealed that these logs include a range of information items, such as *timestamp*, *user ID*, *modeling command*, and *element category*—details also present in the Revit native log. However, custom logs go a step further, capturing additional information items such as *element ID* and *element attribute value changes*. With these additions, chronological changes in building elements can be more thoroughly recorded.

Despite the varied development methods and software employed, such as the Revit C# API, Dynamo, and Rhino API, custom logs differentiate themselves from native logs by incorporating model-element-specific information. Nevertheless, they are not without limitations. Issues of limited public access and the necessity for manual updates could curb the effectiveness of BIM log mining.

The lexicon-based content analysis of native logs from the major BIM authoring tools revealed disparities in content. Whereas Allplan logs contain only *timestamp* and *session ID*, Archicad logs exclude *timestamp*, *element ID*, and *element attribute value change*. On the other hand, Revit and Vectorworks logs cover most information items. However, all of these native logs uniformly lack *element ID* and *element attribute value change*, both of which are deemed crucial for comprehending the BIM process.

Supportability of BIM logs for use cases: Supportability of BIM logs for use cases emerges from two key observations. First, the supportability of a log can be influenced by the absence of process mining prerequisites (i.e., timestamp, case ID, or activity). For example, Allplan's native log fails to support any use case owing to the lack of information items that can represent activities. Archicad's native log contains the *user ID* and *modeling command*, but only supports one use case (CD1) owing to its lack of a timestamp.

Second, the availability of model-element-specific information, such as *element ID* and *element attribute value changes*, significantly influences the supportability of a log for advanced BIM log mining applications. In this regard, custom logs typically outperform native logs, with the latter supporting merely 8 of the 19 use cases. Interestingly, native logs predominantly support process discovery use cases, yet show limited support for process enhancement and conformance checking, which are considered as mature applications of process mining. This lack of support extends to use cases such as CD4, CM1, CC1, and PP1, all of which involve BIM log mining for construction management, and are not supported by any native log owing to the absence of model-element-specific information.

Based on these analyses, the future directions for BIM log mining could be as follows:

Expanding use cases: BIM log mining research should explore a wider variety of process mining applications, such as replaying historical data and generating recommendations. This could lead to further refinement of existing techniques and the development of new ones, particularly with the increased incorporation of AI-driven improvements and automated processes.

Extending the focus to the entier project lifecycle: Currently, BIM log mining is project-specific, focusing on individual project performances and developments of specific building elements, mainly on design phases. By expanding the focus to encompass the entire project lifecycle, a more holistic and comprehensive understanding of the process could be obtained, broadening the scope and potential of BIM log mining.

Enhancing native BIM logs: The role of vendors is crucial in the progression of BIM log mining. The authors of this study encourage them to enhance native BIM logs, not only to include more comprehensive information related to user behavior such as user ID and specific actions users take, but also to incorporate model-element-specific information items. This would enable a more reliable and stable recording of chronological changes of building elements, thereby broadening the scope of BIM log mining and significantly improving its supportability for advanced applications.

Utilizing BIM logs in practice: Industry practitioners should actively collect and analyze BIM logs to optimize their processes, increase efficiency, and improve productivity. BIM log mining can provide insights into participant behavior, element development, and overall project performance that can inform strategic decision-making and operational improvements.

The future of BIM log mining will include integrating more advanced techniques, broadening the scope of analysis, and improving the quality and utility of BIM logs, all while maintaining a strong focus on practical, industry-relevant applications.

8. Conclusion

This paper analyzes the information requirements for various BIM log mining use cases and the extent to which the logs of major BIM authoring tools and custom-developed BIM loggers support each BIM log mining use case to investigate the following research topics:

(1) **Use cases of BIM log mining and their information requirements:** This study investigated various use cases of BIM log mining from existing literature, determining that process discovery was predominant and process enhancement was an emerging use case. Each use case displayed distinct information requirements. These involved process mining prerequisites like timestamps, activities, and case ID, but also incorporated unique nature tied specifically to AECO projects, focusing on user-based analysis and building element-based analysis.
(2) **Contents of BIM logs:** Through a lexicon-based content analysis of BIM logs recorded using major BIM authoring tools and custom-developed BIM loggers, a list of information items that could be collected via BIM logs of commercial BIM authoring tools was analyzed. The items that could be obtained from BIM logs were *timestamp*, *user ID*, *modeling command*, and *element category*. Custom logs generally exceeded native logs in their capacity to support a broad spectrum of log items, such as *element ID* and *element attribute value changes*.
(3) **Supportability of BIM logs for various use cases:** By comparing the information required by various use cases of BIM logs and the contents of BIM logs recorded in major BIM authoring tools, the supportability of various use cases using the major BIM authoring tools was analyzed. The analysis determined that native BIM loggers should be further developed to support model-element-specific information items, which would allow for the expansion of BIM log mining to a model-element level using commercial BIM authoring tools.

This study contributes to BIM log mining in both academic and practical contexts and elucidates possible future directions. From an academic perspective, it identifies potential areas of future research in BIM log mining, with a particular focus on the possibility of applying emerging data analysis techniques to expand use cases. It also suggests investigating ways to extend the lifecycle focus of BIM log mining and refine the mining techniques for more detailed and project-specific outcomes. For industry practitioners, these findings provide useful guidance for improving the effectiveness of BIM log mining in practice. Active collection and analysis of BIM logs can lead to significant enhancements in process optimization, efficiency, and productivity, as well as increased understanding of one's own processes. Moreover, the study calls upon vendors to enhance native BIM logs by including comprehensive user behavior and model-element-specific information, which would help broaden the scope of BIM log mining.

In conclusion, this study investigated the current state of BIM log mining, examining the use cases, content, and supportability of BIM logs for diverse use cases. Through its focus on user and model-element-specific analysis, BIM log mining differentiates itself from traditional process mining, revealing unique requirements that both academic researchers and industry practitioners must consider. The research identifies areas for enhancement in BIM log mining, primarily through the inclusion of comprehensive user behavior and model-element-specific data in BIM logs. These enhancements could potentially improve BIM process representation, help to optimize and automate design processes, and expand the applications of BIM log mining. The exact impact of these enhancements, however, remains to be determined.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

This work is supported in 2023 by the Korea Agency for Infrastructure Technology Advancement(KAIA) grant funded by the Ministry of Land, Infrastructure and Transport (Grant RS-2021-KA163269).

References

[1] A.Z. Onur, F. Nouban, Software in the architectural presentation and design of buildings: state-of-the-artInternational, J. Innov. Technol. Explor. Eng. (IJITEE) [Online] 8 (2019).

[2] S. Yarmohammadi, R. Pourabolghasem, A. Shirazi, B. Ashuri, A sequential pattern mining approach to extract information from BIM design log files, in: ISARC 2016 - 33rd International Symposium on Automation and Robotics in Construction, 2016, pp. 174–181. https://doi.org/10.22260/isarc2016/0022.

[3] S. Yarmohammadi, R. Pourabolghasem, D. Castro-Lacouture, Mining implicit 3D modeling patterns from unstructured temporal BIM log text data, Autom. Constr. 81 (2017) 17–24, https://doi.org/10.1016/j.autcon.2017.04.012.

[4] Y. Pan, L. Zhang, M.J. Skibniewski, Clustering of designers based on building information modeling event logs, Comput.-Aid. Civ. Infrastruct. Eng. 35 (2020) 701–718, https://doi.org/10.1111/mice.12551.

[5] Y. Pan, L. Zhang, Z. Li, Mining event logs for knowledge discovery based on adaptive efficient fuzzy Kohonen clustering network, Knowl.-Based Syst. 209 (2020), 106482, https://doi.org/10.1016/j.knosys.2020.106482.

[6] T. Ishizawa, Keystone players in collaborative building information modeling—form of contribution in Japanese large-scale projects, J. Inf. Technol. Constr. 27 (2022) 685–704. https://doi.org/10.36680/j.itcon.2022.034.

[7] L. Zhang, B. Ashuri, BIM log mining: discovering social networks, Autom. Constr. 91 (2018) 31–43, https://doi.org/10.1016/j.autcon.2018.03.009.

[8] S. Kouhestani, M. Nik-Bakht, Towards level 3 BIM process maps with IFC & XES process mining, EWork and EBusiness in Architecture, Engineering and Construction - Proceedings of the 12th European Conference on Product and Process Modelling, ECPPM 2018, 2018, pp. 103–111. https://doi.org/10.1201/9780429506215-14.

[9] S. Kouhestani, M. Nik-Bakht, IFC-based process mining for design authoring, Autom. Constr. 112 (2020), 103069, https://doi.org/10.1016/j.autcon.2019.103069.

[10] Y. Pan, L. Zhang, A BIM-data mining integrated digital twin framework for advanced project management, Autom. Constr. 124 (2021), https://doi.org/10.1016/j.autcon.2021.103564.

[11] Y. Pan, L. Zhang, Sequential design command prediction using BIM event logs, in: Construction Research Congress 2020: Computer Applications, American Society of Civil Engineers, Reston, VA, 2020, pp. 306–315.

[12] Y. Pan, L. Zhang, BIM log mining: Learning and predicting design commands, Autom. Constr. 112 (2020), 103107, https://doi.org/10.1016/j.autcon.2020.103107.

[13] W. Gao, X. Zhang, Q. He, B. Lin, W. Huang, Command prediction based on early 3D modeling design logs by deep neural networks, Autom. Constr. 133 (2022), https://doi.org/10.1016/j.autcon.2021.104026.

[14] W. van der Aalst, A. Adriansyah, A.K.A. de Medeiros, F. Arcieri, T. Baier, T. Blickle, J.C. Bose, P. van den Brand, R. Brandtjen, J. Buijs, Process mining manifesto, in: International Conference on Business Process Management, Springer, 2011, pp. 169–194.

[15] R.J.C. Bose, R.S. Mans, W.M. van der Aalst, Wanna improve process mining results?, in: IEEE Symposium on Computational Intelligence and Data Mining (CIDM) IEEE, 2013, pp. 127–134, https://doi.org/10.1109/CIDM.2013.6597227.

[16] S. Suriadi, R. Andrews, A.H. ter Hofstede, M.T. Wynn, Event log imperfection patterns for process mining: towards a systematic approach to cleaning event logs, Inf. Syst. 64 (2017) 132–150.

[17] A. Rozinat, W.M.P. van der Aalst, Conformance checking of processes based on monitoring real behavior, Inf. Syst. 33 (2008) 64–95, https://doi.org/10.1016/j.is.2007.07.001.

[18] M.L. van Eck, X. Lu, S.J. Leemans, W.m., Van Der Aalst, PM^2: a process mining project methodology, in: International Conference on Advanced Information Systems Engineering, Springer, 2015, pp. 297–313.

[19] S. Yarmohammadi, D. Castro-Lacouture, Automated performance measurement for 3D building modeling decisions, Autom. Constr. 93 (2018) 91–111, https://doi.org/10.1016/j.autcon.2018.05.011.

[20] W. Gao, C. Wu, W. Huang, B. Lin, X. Su, A data structure for studying 3D modeling design behavior based on event logs, Autom. Constr. 132 (2021), https://doi.org/10.1016/j.autcon.2021.103967.

[21] S. Kannan, S. Karuppusamy, A. Nedunchezhian, P. Venkateshan, P. Wang, N. Bojja, A. Kejariwal, Chapter 3 - big data analytics for social media, in: R. Buyya, R. N. Calheiros, A.V. Dastjerdi (Eds.), Big Data, Morgan Kaufmann, 2016, pp. 63–94, https://doi.org/10.1016/B978-0-12-805394-2.00003-9.

[22] Y. Pan, L. Zhang, BIM log mining: exploring design productivity characteristics, Autom. Constr. 109 (2020), 102997, https://doi.org/10.1016/j.autcon.2019.102997.

[23] E. Forcael, A. Martinez-Rocamora, J. Sepulveda-Morales, R. Garcia-Alvarado, A. Nope-Bernal, F. Leighton, Behavior and performance of bim users in a collaborative work environment, Appl. Sci.-Basel. 10 (2020) 2199, https://doi.org/10.3390/app10062199.

[24] Y. Pan, L. Zhang, Roles of artificial intelligence in construction engineering and management: a critical review and future trends, Autom. Constr. 122 (2021), 103517, https://doi.org/10.1016/j.autcon.2020.103517.

[25] Y. Pan, L. Zhang, Modeling and analyzing dynamic social networks for behavioral pattern discovery in collaborative design, Adv. Eng. Inform. 54 (2022), 101758, https://doi.org/10.1016/j.aei.2022.101758.

[26] Y. Pan, L. Zhang, Automated process discovery from event logs in BIM construction projects, Autom. Constr. 127 (2021), 103713, https://doi.org/10.1016/j.autcon.2021.103713.

[27] L. Zhang, M. Wen, B. Ashuri, BIM log mining: measuring design productivity, J. Comput. Civil. Eng. 32 (2018) 04017071, https://doi.org/10.1061/(ASCE)CP.1943-5487.0000721.

[28] B. Zhong, H. Wu, L. Ding, P.E. Love, H. Li, H. Luo, L. Jiao, Mapping computer vision research in construction: developments, knowledge gaps and implications for research, Autom. Constr. 107 (2019), 102919.

[29] W. Gao, S. Lu, X. Zhang, Q. He, W. Huang, B. Lin, Impact of 3D modeling behavior patterns on the creativity of sustainable building design through process mining, Autom. Constr. 150 (2023), 104804, https://doi.org/10.1016/j.autcon.2023.104804.

[30] W. Van Der Aalst, Process Mining: Data Science in Action, Springer, Berlin, Heidelberg, 2016.

[31] C. dos S. Garcia, A. Meincheim, E.R. Faria Junior, M.R. Dallagassa, D.M.V. Sato, D. R. Carvalho, E.A.P. Santos, E.E. Scalabrin, Process mining techniques and applications – a systematic mapping study, Expert Syst. Appl. 133 (2019) 260–295, https://doi.org/10.1016/j.eswa.2019.05.003.

[32] K. Krippendorff, Content Analysis: An Introduction to Its Methodology, Sage Publications, 2018.

[33] P. Mayring, Qualitative content analysis, Compan. Qual. Res. 1 (2004) 159–176.

[34] K.A. Neuendorf, The Content Analysis Guidebook, Sage Publications, 2017.

[35] R. Sacks, C. Eastman, G. Lee, P. Teicholz, BIM Handbook: A Guide to Building Information Modeling for Owners, Designers, Engineers, Contractors, and Facility Managers, John Wiley & Sons, 2018.

[36] Autodesk, About Journal Files, Autodesk REVIT 2022 Help, 2022. https://help.autodesk.com/view/RVT/2022/ENU/?guid=GUID-477C6854-2724-4B5D-8B95-9657B636C48D (Accessed November 22, 2022).

[37] S. Jang, S. Shin, G. Lee, Logging Modeling Events to Enhance the Reproducibility of a Modeling Process, in: ISARC Proceedings of the International Symposium on Automation and Robotics in Construction, IAARC Publications, 2021, pp. 256–263.