Journal of Engineering, Project, and Production Management
2024, 14(4), 0036

A Knowledge-Driven Approach to Automate Job Hazard Analysis Process

Sonali Pandithawatta[1], Raufdeen Rameezdeen[2], Seungjun Ahn[3], Christopher W. K. Chow[4], and Nima Gorjian[5]

[1]Ph.D. Student, UniSA STEM, University of South Australia, Adelaide, SA 5000, Australia, E-mail: thalpe_panditha_wattage_sonali.pandithawatta@mymail.unisa.edu.au (corresponding author).
[2]Professor, UniSA STEM, University of South Australia, Adelaide, SA 5000, Australia, E-mail: rameez.rameezdeen@unisa.edu.au
[3]Assistant Professor, Department of Civil and Environmental Engineering, Hongik University, Seoul 04066, Republic of Korea, E-mail: jun.ahn@hongik.ac.kr
[4]Professor, UniSA STEM, University of South Australia, Adelaide, SA 5000, Australia, E-mail: christopher.chow@unisa.edu.au
[5]Associate Professor, UniSA STEM, University of South Australia, Adelaide, SA 5000, Australia, E-mail: nima.gorjianJolfaei@unisa.edu.au

Project Management
Received November 24, 2023; received revision November 21, 2024; accepted November 21, 2024
Available online November 28, 2024

Abstract: Automating the job hazard analysis (JHA) process is an urgent requirement in the construction safety management field due to limitations of the conventional process. The manual nature of conducting the JHA and the dynamic environment of construction sites make it necessary to perform the analysis before commencing the job and to then regularly update it in accordance with changes in the construction plans. With this in mind, this research aims to develop an automated approach to support safety personnel during the JHA process.

In seeking to automate the JHA process, the nature of construction accidents, hazards and risk assessment needs to be studied in light of the theoretical knowledge on accident causation. Thus, this research was designed according to the constructive research approach to develop a job hazard analysis knowledge graph (JHAKG) to automate the JHA process. The JHAKG incorporated an ontology (O-JHAKG) built according to the systematic ontology development method, METHONTOLOGY, which formalises both explicit and implicit knowledge inherent in the JHA process. The data were imported to the JHAKG from an incident database using rule-based natural language processing (NLP) which helped to extract implicit information not evident in the traditional JHA document. The validation of the JHAKG was conducted in two stages: the first stage validated the information extraction process by calculating performance metrics, while the second stage validated the data population process and the JHAKG's reasoning capability. The overall research resulted in a comprehensive JHAKG with advanced inferencing capabilities which can assist safety personnel in effectively executing the JHA process.

Keywords: Construction industry, knowledge graph, nlp, ontology, safety management.

DOI 10.32738/JEPPM-2024-0036

---

1. Introduction

The construction industry is widely acknowledged as a high-risk industry with one of the most unsafe working environments, with approximately one in five occupational fatalities occurring in this industry (Asadzadeh et al., 2020). This can be attributed to various characteristics of the construction process, such as extended construction duration, unique site conditions, multifaceted construction techniques and a strong dependence on environmental factors (Moghadami and Mortazavi, 2018). Consequently, a rise in construction activities may lead to a corresponding increase in occupational accidents and fatalities. Therefore, it is imperative to prioritise workers' safety by fostering intensified safety awareness and implementing proactive accident prevention measures (Xing et al., 2019).

Job hazard analysis (JHA) is recognised as a systematic approach to mitigate safety risks and is regarded as a fundamental element of the safety management system (Crutchfield and Roughton, 2019; Glenn, 2011). Serving as a planning tool, JHA enhances safety, productivity and the overall quality of the tasks to be completed (Albrechtsen, Solberg, and Svensli, 2019). Those responsible for conducting JHA rely on collaborative brainstorming sessions to identify the sequential steps involved in various construction activities and to recognise the associated hazards. During JHA, they draw from their experience and commonly utilise safety knowledge, represented in the form of safety rules (Wang and Boukamp, 2011). Additionally, previous JHA documents and incident reports are used as guidance. Therefore, JHA can be characterised as a knowledge-intensive activity. Consequently, the JHA execution has become increasingly challenging, prompting researchers to seek solutions through the formalisation of knowledge.

Goh and Chua (2009, 2010) developed an approach based on case-based reasoning to identify construction safety hazards. The objective of their approach is to leverage prior knowledge, derived from past hazard identification and incident cases, to improve the effectiveness and quality of new hazard identification. Wang and Boukamp (2011) developed a framework with the objective of enhancing access to a company's JHA knowledge by streamlining the complexity and time-consuming nature of the traditional JHA process. The framework utilises ontologies to organise knowledge related to activities, job steps and hazards. Furthermore, it incorporates an ontological reasoning mechanism to identify safety rules that are relevant to specific activities. Chi, Lin and Hsieh (2014) identified the potential of using existing construction safety resources to assist JHA, with the goal of reducing the level of human effort required. The authors employed ontology-based text classification to match safe approaches found in existing resources with unsafe scenarios. Their study's findings offer valuable support for carrying out the JHA process by automatically retrieving relevant safety documents for various situations. Zhang, Boukamp and Teizer (2015) devised a construction safety ontology to formalise construction safety management knowledge. Their ontology encompasses a construction product model, a construction process model and a construction safety model. It is seamlessly integrated with the Building Information Modelling (BIM) platform to visually represent inferred knowledge, including the necessary protective safety systems and zones. This integration facilitates and supports automated ontology-based JHA processes. Ding, Zhong, Wu and Luo (2016) introduced a framework based on ontology, semantic web technology and BIM to automatically establish relationships between construction risk factors, causes and preventive measures. The researchers also developed a prototype system as a tool to enhance the management and reuse of construction risk knowledge.

Implicit knowledge, which is closely tied to domain experts' experience, is a critical component in the precise and comprehensive execution of the JHA process (Altawil, 2017; Pandithawatta et al., 2023). However, most prior studies have not implemented a systematic method to extract implicit knowledge to understand the underlying reasons behind critical JHA decisions. During the knowledge formalisation process, these studies were thus unable to comprehensively transform implicit knowledge into computer-accessible format. As a result, automated JHA systems often have less robust reasoning capabilities, tending to generate generic information about hazards and preventive measures. Considering these limitations, this research is focused on building a knowledge graph (JHAKG) to support JHA by integrating the implicit knowledge of domain experts. Furthermore, a rule-based natural language processing (NLP) information extraction mechanism is implemented to extract information from an incident database, enabling the developed JHAKG to be populated. This allows users to access both previous JHA information and incident data simultaneously while performing the JHA process.

2. Method

2.1. Constructive Research Approach

The constructive research approach aims to solve real-world problems by producing innovative constructions while contributing to the theory of the domain in which it is applied (Crnkovic, 2010). The word "constructions" refers to a wide range of human artefacts, such as diagrams, models, information systems, algorithms, plans, organisation structures and designed systems (Chen, Lu, Fu, and Dong, 2023; Lukka, 2000). The constructive research approach is similar to other case/field research approaches, such as ethnographic research, grounded theory, illustrative case research, theory testing case research and action research (Lukka, 2000). Constructive research and action research are extensively used in developing and implementing novel computing and information technology (IT) approaches across various domains (McGregor, 2018). However, it is important to note that constructive research is distinct from action research in two ways. Firstly, constructive research always focuses on the construct as an outcome, whereas action research may have different goals. Secondly, the interaction of researchers with practice and practitioners is not mandatory in constructive research, whereas it is necessary in action research (Lehtiranta, Junnonen, Kärnä, and Pekuri, 2015). The constructive research approach can be presented as a type of applied study that produces new knowledge as a normative application (Kasanen, Lukka, and Siitonen, 1993). Results from the constructive research approach should show how to act in a situation to achieve a specific outcome. The assumption is that if a certain action is taken, it will lead to a particular result. Without the assumption about the causality of things, presenting technical norms of this kind would be illogical. The normative nature of constructive research and its focus on bringing about real-world change set it apart from other types of research (Lehtiranta et al., 2015).

2.2. Rationale for Selecting the Constructive Research Approach

In seeking to produce an automated JHA system, it is important to follow a research approach which facilitates the development of a new human artefact. Given the nature of this research, the constructive research approach was identified as the most suitable. Table 1 presents the core features of the constructive research approach, as explained by Lukka (2000) and how they are matched with the features of this research.

Table 1. Applicability of constructive research approach

No	Features of constructive research approach	Nature of the research
1	Focuses on addressing practical real-world problems	The research problem of this study is the practical issue of the inefficient, time-consuming and labour-intensive nature of the JHA process in the construction industry, with this having received much attention from practitioners and researchers.
2	Creates an innovative construct meant to solve the initial managerial problem	This research focuses on producing a job hazard analysis knowledge graph (JHAKG) to assist construction professionals to execute the JHA process in a more effective and comprehensive manner.
3	Includes an attempt to implement the developed construct and thereby test its practical applicability	The research process includes proposed steps for implementing and validating the developed JHAKG which are intended to test its operational ability.
4	Implies a collaborative approach between the researcher and practitioner in a team-like manner, with a focus on experiential learning	The researchers worked closely with construction safety experts to gain their ideas, experience, knowledge and feedback, ensuring that the research findings would be relevant and applicable to real-world situations.
5	Is explicitly linked to previous theoretical knowledge	The researchers bring prior theoretical knowledge on accident causation into the research process, allowing them to analyse hazard information to construct a knowledge graph schema for the JHA process.
6	Pays special emphasis to relating the empirical findings back to theory	The JHAKG, which the researchers construct, underpins a new body of knowledge that can be studied and understood. Therefore, the knowledge embedded in the JHAKG has undeniable epistemological value, providing a unique opportunity for researchers to gain new insights and knowledge.

As shown in Table 1, the constructive research approach features are clearly well matched with the nature of the current research study. This approach not only lays solid groundwork for developing an artefact but also offers researchers the means to accomplish their objectives. Therefore, this research adopts the constructive research approach to develop a knowledge graph (i.e., the JHAKG) that automates the JHA process in the construction industry. The detailed development process is discussed in the following section.

3. Job Hazard Analysis Knowledge Graph (JHAKG) Development Process

When an engineer or scientist creates something, such as an artefact, following the constructive research approach, it is important that the construction process is studied and analysed with the same level of rigour used in other research methods, like grounded theory and action research (Crnkovic, 2010). Researchers should focus not only on the end result or product, but also on its construction process. By studying the construction process in detail, researchers can gain a better understanding of the underlying principles and methods used to build the artefact. This understanding can help improve the quality and efficiency of the construction process, leading to better and more effective products in the future. Given these considerations, the construction of the JHAKG was systematically conducted by following well-established methods. The development process of the JHAKG, as presented in Fig. 1, consists mainly of two phases: (i) the development of the ontology of the JHAKG (O-JHAKG) and (ii) the O-JHAKG population process. The following sub-sections elaborate on the current research study's ontology development and data population methods, with reasons provided for their selection.

3.1. O-JHAKG Development

As highlighted in the introduction, to improve the performance of an automated system, it is vital to incorporate the implicit knowledge of safety experts. The reason is that this knowledge has a strong impact on the system's reasoning capability. Incorporating this implicit knowledge requires a manual ontology development method. In their study, D'Avanzo, Lieto and Kufik (2008) found that manual ontology construction facilitates the insertion of meaningful information into the ontological system. With the direction of human expertise, this leads to the creation of the taxonomy's concepts. Therefore, a comprehensive manual ontology development method that can lead and manage the development process is critical for the proposed JHAKG's quality. The current research thus utilised METHONTOLOGY, one of the leading manual ontology engineering methods (Fernández-López, Gómez-Pérez, and Juristo, 1997). As a well-structured methodology, METHONTOLOGY is utilised to build ontologies from scratch, with its ontology development steps transparent and logically complete.

An ontology development methodology should encompass a set of well-established principles, processes, practices, methods and activities, all of which are used to design, construct, evaluate and deploy ontologies (Gasevic, Djuric, and Devedzic, 2009). Incorporating these essential components, METHONTOLOGY is identified as a comprehensive methodology for building ontologies (Elmhadhbi, Mohamed-Hedi, Archimède, Otte, and Smith, 2021). The six phases of METHONTOLOGY start with (1) the first phase, specification, which involves defining the purpose, scope, domain and requirements of the ontology with the help of competency questions. The second phase, (2) knowledge acquisition, focuses on gathering knowledge from experts, written sources, figures and previous ontologies. Knowledge acquisition techniques, such as brainstorming, interviews and text analysis, are utilised in this phase, beginning in the specification phase and reducing as the ontology development process progresses. In the third phase, (3) conceptualisation, the domain knowledge is structured into a conceptual model using a glossary of terms that represent the domain knowledge and its meanings. The fourth phase, (4) integration, aims to expedite the ontology development process by integrating other compatible ontologies instead of starting from scratch. The fifth phase, (5) implementation, requires an ontology development environment that supports the meta-ontology. This phase results in a codified ontology written in a formal ontology language. Finally, the sixth phase, (6) evaluation, also known as the verification and validation phase, involves the technical evaluation of the developed ontology (Fernández-López et al., 1997).

The general METHONTOLOGY process was adjusted logically according to this research's requirements. Although the specification and knowledge acquisition phases were conducted simultaneously, to facilitate the presentation, the discussion begins with knowledge acquisition followed by specification. Even though no other ontologies were integrated into the O-JHAKG, previous ontologies related to the JHA provided a basic understanding of the broad concepts and associations relevant to the JHA knowledge domain. Therefore, the integration phase was included in the conceptualisation phase to represent this step. The detailed discussion of each of these phases is presented in the following sub-sections.

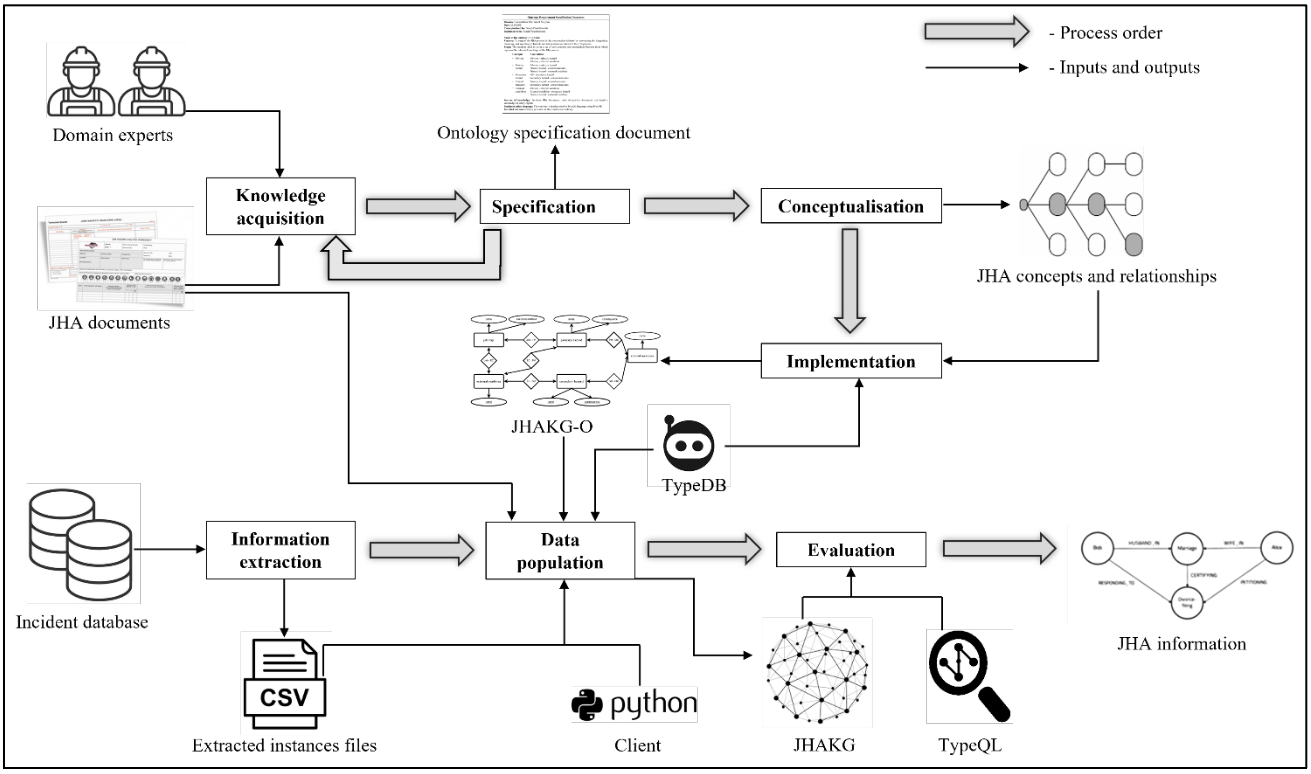


Fig. 1. Job hazard analysis knowledge graph (JHAKG) development process

3.1.1. Knowledge acquisition

Developing an ontology requires a comprehensive understanding of the targeted domain. Therefore, knowledge needs to be acquired from the available sources to gain an overall idea about the domain. As the effectiveness of a comprehensive JHA is not solely influenced by explicit knowledge sources, safety personnel's implicit knowledge acquired through everyday experiences should be included in an ontology representing JHA knowledge. Those who possess implicit knowledge often find it difficult to articulate this knowledge or even to realise it (Rosson and Carroll, 2002; Sternberg et al., 2000). Thus, the current research's method was structured to enable the acquisition of implicit knowledge from safety personnel to create an ontology with enhanced reasoning capacity.

Knowledge acquisition was conducted in two stages. In the first stage, a thorough document analysis was conducted to identify hazardous activities and hazards, and control measures. In total, 115 JHA documents were collected from water infrastructure contractors, representing various common construction activities and water infrastructure-specific activities. The second stage comprised a Delphi study, with this method having gained significant recognition as an effective approach for systematically exploring the implicit knowledge of experts to establish a consensus within a group on a specific topic or problem (Dayé, 2018; Harteis, 2022; Hillege, van Domburgh, Mulder, Jansen, and Vermeiren, 2018;

Niederberger and Köberich, 2021; Spickermann, Zimmermann, and von der Gracht, 2014). To develop an ontology to represent the JHA knowledge domain, the safety experts' reasoning for their decisions made during the JHA process needed to be made explicit. Moreover, in their studies, Cooke, Lingard, Blismas and Stranieri (2008) and Poghosyan et al. (2020) successfully utilised the Delphi method to facilitate the development of web-based tools as knowledge-driven applications in the field of construction safety management. Therefore, a Delphi study was identified as the most suitable approach for identifying the different concepts and relationships in the targeted domain. Furthermore, the current research chose a qualitative approach to retain the nuanced understanding obtained during the initial narrative brainstorming throughout the entire research process. As a result, to extract the implicit knowledge embedded in the JHA process, a three-round qualitative Delphi study was conducted with 18 participants. Of these 18 participants, 56% were directly engaged in the daily JHA process, while the remaining 44% had an indirect involvement. The participants' length of experience ranged from 5–34 years, with an average of 20 years. The sample encompassed individuals from diverse job positions and included supervisors, safety consultants, site safety advisors and safety managers. Therefore, the sample comprised a diverse group of domain experts with significant experience and expertise in the JHA process.

3.1.2 Specification

Fernández-López et al. (1997) emphasised the importance of establishing the purpose and scope of the ontology before initiating its development. The scope serves as a boundary for the ontology, defining what should be included and what should be excluded. This step plays a crucial role as it streamlines the analysis by minimising the amount of data and concepts to be considered (Brusa, Caliusco, and Chiotti, 2006). Identifying competency questions in the specification phase is vital for defining the scope of the ontology. Competency questions consist of a set of questions which the ontology is capable of answering (Khatoon, Hafeez, and Ali, 2014). During the initial round of the Delphi study, the domain experts explained four main requirements of the automated JHA system. The competency questions, which cover the main concepts of the O-JHAKG, focused on these four main requirements: (1) recognising primary hazards; (2) identifying control measures for primary hazards; (3) recognising secondary hazards and their control measures; and (4) evaluating changes in risk levels due to external conditions.

The ontology requirement specification document outlines the requirements and specification for the development of the ontology. This document is important during the ontology development process to ensure that the ontology meets the needs of its intended users. Thus, it is an important tool to ensure that the resultant ontology is well designed, effective and meets the requirements of its intended audience. According to Fernández-López et al. (1997), the content of a good ontology specification document should be relevant without duplications, partially complete so it can be updated anytime and consistent in terms of meaning so it can be understood within the domain. Fig. 2 below presents the O-JHAKG requirement specification document, describing its purpose, scope and intended end-users.

Ontology Requirement Specification Document

Domain: Construction Job Hazard Analysis
Date: 22.05.2021
Conceptualized by: Sonali Pandithawatta
Implemented by: Sonali Pandithawatta

Name of the ontology: O-JHAKG
Purpose: To support the JHA process in the construction industry by automating the integration, reasoning, and searching of hazards and risk information related to the JHA process.
Scope: The ontology need to cover a set of core concepts and associations between them which represent the inherent knowledge of the JHA process.

Concepts	Associations
• Job step	Job step - primary hazard Job step - external condition
• Primary hazard	Job step - primary hazard Primary hazard - control measures Primary hazard - external condition
• Secondary hazard	Job - secondary hazard Secondary hazard - control measures
• Control measures	Primary hazard - control measures Secondary hazard - control measures
• External conditions	Job step - external condition External condition - secondary hazard Primary hazard - external condition

Sources of knowledge: Previous JHA documents, code of practice documents and implicit knowledge of safety experts
Implementation language: The ontology is implemented in TypeQL language using TypeDB
Intended end users: Safety personnel in the construction industry

Fig. 2. Ontology specification document

3.1.3. Conceptualisation

The conceptualisation phase plays a crucial role in the ontology development process, identifying and defining the concepts and relationships to be incorporated (Fernández-López et al., 1997). METHONTOLOGY recommends the incorporation of the 'integration' step to allow the reuse of existing ontologies which significantly accelerates the ontology development process. Hence, the current research reviewed previous ontologies and meta-models on the identification of hazards and control measures. As these prior studies did not incorporate implicit knowledge, their ontologies did not define the concepts as needed in the O-JHAKG; therefore, it was conceptualised from scratch, with no reliance on previous work.

Interview data and JHA documents were analysed to recognise meaningful concepts utilised by domain experts in the JHA process. These concepts were categorised into entities and attributes, with associations between them determined. As the resulting ontology's quality would be significantly influenced by the precision of the analysis, reasoning ability heavily depended on the integrated entities, attributes and relationships. Therefore, a rigorous data analysis was conducted to ensure a thorough and detailed understanding of the JHA process. An abductive approach was utilised to systematically code and classify the data for analysis. This approach ensured that the analysis was not only guided by the theoretical understanding of the systems model of construction accident causation but was also driven by the collected data (Mitropoulos, Abdelhamid, and Howell, 2005; Naisola-Ruiter, 2022; Thompson, 2022). The Mitropoulos systems model developed by Mitropoulos, Abdelhamid and Howell (2005) considers a systems view of construction accidents by examining how production system features contribute to the emergence of hazardous situations during activities and influence work behaviours. It also analyses the conditions that have an impact on the likelihood of hazards. Thus, this model served as a valuable theoretical foundation for the current research, guiding the analysis of the collected data. The following sub-sections describe the O-JHAKG's main entities along with their attributes that contribute to making inferences in the JHA process. Although the sub-sections do not elaborate on the detailed analysis results of the JHA documents and interview data, they present a summary of the results of the analysis process.

3.1.3.1 Job step

To infer the hazards associated with a job step (identified in the current research as primary hazards), the following combination needed to be defined within the job step: (i) the associated act; (ii) the execution method; and (iii) the associated element. Once a job step was identified as a combination of these attributes, the relationship between the job step and the primary hazard was activated and the relevant primary hazards were retrieved. For example, to identify a primary hazard associated with the task of "A worker excavating a trench with a mini-excavator to lay a pipeline", the job step needed to be inserted into the JHAKG as a combination of attributes: (i) associated act – *excavating*; (ii) execution method – *mini-excavator*; and (iii) associated element – *trench*.

3.1.3.2. External conditions

Job steps are performed under the influence of various external conditions. Consequently, a relationship was established between job steps and external conditions. This relationship allowed external conditions to influence the probability of the occurrence of certain primary hazards. For example, windy conditions could increase the likelihood of a fire hazard. Thus, another relationship needed to be established between primary hazards and external conditions. However, as not all primary hazards had this relationship, external conditions were integrated into the JHAKG as "IF-THEN" rules. These rules would be useful for assessing the risk of primary hazards based on existing external conditions.

External conditions were sub-categorised based on their nature, that is, conditions related to the weather, the workplace, proximity or the atmosphere. During a job step's execution, in addition to primary hazards, hazards could be caused by these external conditions, referred to as weather hazards, workplace hazards, proximity hazards and atmospheric hazards (collectively identified in the current research as secondary hazards). The attribute influencing the inference of secondary hazards was the external condition's name which activated the relationship between the external condition and secondary hazard entities, thereby identifying the relevant secondary hazard.

3.1.3.3. Primary hazards and secondary hazards

To facilitate the inference of suitable control measures to mitigate primary or secondary hazards, these entities needed to be defined by their (i) name and (ii) consequence. As the hazard's consequence would determine the suitability of the control measures, the inference should not be solely determined by the name of the primary or secondary hazard. Once the hazard was defined by name and consequence, the relationship between the primary or secondary hazard and control measures was activated, with the applicable control measures retrieved. For example, if we considered a noise hazard, the JHAKG would generate two different sets of control measures for the hazard: 'High' consequence and 'Low' consequence.

3.1.3.4. Control measures

Control measures applicable during the stages of the JHA process were identified as isolation, engineering controls, administrative controls and personal protective equipment (PPE). It was important to implement these control measures to mitigate both primary and secondary hazards. Thus, relationships were established between the entities of the control measures and primary and secondary hazards. The control measure was identified by name and incorporated into the O-JHAKG as the attribute that influenced the inferences related to control measures.

Table 2 summarises the core findings of the data analysis, as incorporated into the O-JHAKG, comprising the main entities, sub-entities, attributes and relationships. The implementation and evaluation stages required a populated knowledge graph so these were carried out after populating the O-JHAKG and are later comprehensively discussed.

3.2. O-JHAKG Population

During the practical application of the JHA process, incident reports are referred to, with this step systematically identifying potential hazards that may arise at specific job steps, providing valuable insights and lessons learned from previous incidents. These insights can greatly assist in recognising and mitigating similar hazards in the current job hazard analysis (JHA). However, domain experts rarely follow this practice due to the manual nature of executing the JHA, instead referring to incident databases. This limits their ability to effectively incorporate insights from incident cases into the hazard identification process.

Incidents recorded in a database could include near misses, injuries or accidents experienced by workers while on the worksite. Incident reports describe valuable information relevant to the incident, such as the job step being performed by the worker when the incident occurred; losses incurred; weather conditions; proximity; the atmosphere; nature of the workplace; parties involved; and the incident's severity level. These reports would thus contain a variety of interconnected work health and safety (WHS) information relevant to the safe execution of construction works. The information in the incident reports could be represented as nodes and edges: "job step" – *performed under*→ "external condition" – *created*→ "incident" – *had*→ "severity". These pieces of information, as nodes and edges, were integrated into the JHAKG as they represented the JHA's concepts and relationships. As manual extraction of data into the JHAKG was time consuming and exhausting, a rule-based NLP approach was employed to automatically extract this information from incident reports, thus achieving the JHAKG's population. Through the integration of incident data, this approach facilitated the automation of the JHA, improving comprehensiveness, reliability and ease of execution, while overcoming the traditional process's limitations.

Table 2. Entities, attributes and relationships of the proposed O-JHAKG

Entities	Sub-entities	Attributes	Relationships
Job step	Not applicable	Associated act Associated element Execution method	Job step – primary hazard Job step – external condition
Primary hazard	Not applicable	Name of primary hazard Consequence of primary hazard	Job step – primary hazard Primary hazard – control measures Primary hazard – external condition
Secondary hazard	Weather hazards Workplace hazards Atmospheric hazards Proximity hazards	Name of secondary hazard Consequence of secondary hazard	External condition – secondary hazard Secondary hazard – control measures
Control measures	Isolation control measures Engineering control measures Administrative control measures PPE control measures	Name of control measures	Primary hazard – control measures Secondary hazard – control measures
External conditions	Weather conditions Workplace conditions Atmospheric conditions Proximity conditions	Name of external condition	Job step – external condition External condition – secondary hazard Primary hazard – external condition

For the current study, an incident database covering a period of five years, containing records of near misses, injuries and accidents related to water infrastructure works, was selected as the case study. The rule-based information extraction process was implemented in Python, utilising the SpaCy, Pandas and Numpy packages, and involved five main steps: (i) data pre-processing; (ii) incident extraction; (iii) clause segmentation; (iv) instance extraction; and (v) validation of information extraction. The following sub-sections describe the process in each of these steps.

3.2.1. Data pre-processing

In this step, unwanted data fields, duplications and incidents with descriptions spanning multiple sentences were removed from the database. Subsequently, various pre-processing techniques, including tokenization, part of speech (POS) tagging and noun chunking, were applied to the remaining data to facilitate further linguistic analysis.

3.2.2. Incident extraction

Not all incidents included in the database represented the targeted concepts and the relationships between them. Thus, a sample of the database was manually analysed to identify sentence patterns that exhibited these concepts and their relationships. For instance, incident descriptions such as "*The worker was not feeling well in the morning and left without informing anyone*" did not provide valuable instances for the JHAKG as it did not include a job step and, thus, it lacked a semantic relationship with the incident. However, incident descriptions like "*While welding a steel valve in a manhole, an employee encountered breathing difficulties*" demonstrated the targeted domain knowledge entities (DKEs) and their relationships. The phrase "*welding a steel valve*" represented the job step; "*breathing difficulties*" represented the incident; "*in a manhole*" indicated the nature of the external condition; and "*while*" signified the relationship between the job step and the incident. After identifying sentence patterns that indicated DKEs and the semantic relationships between job steps and incidents, their linguistic features were used to extract compatible incidents from the database by developing matching rules.

3.2.3. Clause segmentation

All extracted incidents from the previous step were compiled into a single Microsoft (MS) Excel file. Each sentence was then segmented into two clauses: the incident clause and the job step clause using the linguistic features of each sentence pattern. The phrase indicating job step information was categorised as the job step clause, while the phrase indicating incident information was categorised as the incident clause. This segmentation was necessary to facilitate precise instance extraction in the subsequent step.

3.2.4. Information extraction

Even though the incidents extracted from the database represented different linguistic structures, compiling the sentences into two distinct segments removed differences in linguistic structures and facilitated the application of common matching rules to each incident for its extraction. Different rules were developed for each concept to precisely extract instances. A token tagged as "VBG" (i.e., a verb, gerund or present participle) represented the job step, while a word or chunk of words tagged as "NN" (i.e., a noun, singular or mass) denoted the associated element. This meant that a dictionary [{"TAG": "VBG"},{"TAG": "NN"}] denoted the job step concept as it captured both the task and its associated element. Similar dictionaries were also developed for the concepts of external condition and executing method, utilising their inherent linguistic features, so instances could be extracted into the JHAKG through the syntactic rules established.

Extracting hazards was a challenging task as the incident description only indicated the accident, injury or near miss without explicitly mentioning the hazard/s that existed prior to the incident. Thus, keyword dictionaries needed to be built to automate the extraction of hazards from the incident clause. Using Australian Codes of Practice documents and WordNet (18), dictionaries of injuries were built. The Codes of Practice documents provided information for identifying hazards from the incident clauses, while WordNet was used to recognise synonyms. Rules were developed to analyse the incident clause, searching for synonyms defined in the keyword dictionaries. If any of the defined synonyms were found in the incident clause, it would indicate the relevant hazard. Some hazards required the use of two or more keyword dictionaries for precise hazard extraction from the incident clause. This was particularly common when the incident was described in more generically rather than in a specific form. For example, incidents resulting from both falling hazards and falling object hazards often included the generic keyword 'fell' in the incident clause, making it difficult to precisely extract the relevant hazard. In these situations, multiple dictionaries were developed to aid in extracting the correct hazard from the incident.

The database adopted a categorisation to indicate the severity level of the incidents as Type 1, Type 2 and Type 3. Severity levels of incidents were indicated as follows: Type 1 'High'; Type 2 'Medium'; and Type 3 'Low'. A hazard's consequence level was determined by considering the severity of the potential incident when its probability reached 100%. The hazard's consequence level was therefore equal to the incident's severity level. Based on this logic, conditional statements were created to extract the consequence level of hazards from the incident reports. At the end of this series of steps, instances that fell under the concepts of job step, hazard, consequence, executing method and external condition could be extracted and the JHAKG could be populated. The instances under control measures were manually extracted from the available JHA documents and Codes of Practice documents, as incident reports did not include these instances.

Table 3. Measures for performance evaluation

Metrics	Description
True positive (TP)	Number of precisely extracted instances
False positive (FP)	Number of incorrectly extracted instances
False negative (FN)	Number of non-extracted precise instances
Precision	$\frac{TP}{TP+FP}$
Recall	$\frac{TP}{TP+FN}$
F-measure $0 \leq \alpha \leq 1$	$\frac{Precision\ X\ Recall}{(1-\alpha)\times precision+\alpha\times recall}$

Table 4. Precision, recall and F-measure for concepts

Metrics \ Concepts	Job step	Hazard	Consequence	Execution method	External condition
Gold std.	150	150	150	37	77
TP	150	135	150	35	77
FP	0	7	0	4	3
FN	0	15	0	2	0
Precision	1.000	0.951	1.000	0.897	0.963
Recall	1.000	0.900	1.000	0.946	1.000
F-measure	1.000	0.925	1.000	0.921	0.981

3.2.5. Validation of information extraction

The performance of the information extraction mechanism was validated against a manually created gold standard file which consisted of instances annotated by domain experts that needed to be precisely extracted using NLP algorithms. Five safety experts with 10–20 years of experience in water infrastructure works participated in the process of creating the gold standard, with 150 randomly selected incident reports used for this purpose. Each expert received 30 incident reports and was asked to extract the job step, hazard, consequence and external condition from each incident report and to indicate if this information was implicitly or explicitly mentioned. The resulting gold standard was compared to the results generated from the NLP information extraction mechanism adopted in the current research, and the performance evaluation metrics (precision, recall and F-measure) were calculated, as denoted in Tables 3 and 4.

3.3. Implementation

To build the O-JHAKG, the schema was loaded into TypeDB, a deductive database designed for artificial intelligence applications (Vaticle, 2023). TypeDB consists of two major components: the TypeDB database, responsible for storing and managing knowledge, and the TypeQL query language, used for interacting with the database and executing complex queries. The instances of the gold standard and preventive measures, extracted manually from previous JHAs, were saved in the form of comma-separated values (CSV) files. These files were then migrated to the O-JHAKG, which was running on the TypeDB server, utilising a Python client. During this step, each data item was parsed into a Python dictionary and a TypeQL query was constructed to insert the data item into the O-JHAKG, leading to the JHAKG's effective creation. This process ensured that the data from the CSV files were successfully imported into the O-JHAKG in TypeDB, enabling efficient storage, retrieval and querying of JHA-related information.

3.4. JHAKG Evaluation

After creating the JHAKG, the successful execution of the data migration process needed to be verified to ensure that the resulting knowledge graph could generate accurate responses to queries. For this verification, the gold standard was utilised as the data included in it had already been verified by domain experts. Thus, it was only necessary to assess whether the JHAKG produced correct responses to queries related to the gold standard. To accomplish this, the competency questions formulated during the specification stage were employed. Some of these questions were related to identifying hazards associated with job steps; determining the control measures to be implemented during job steps; and evaluating the risk levels of hazards in relation to external conditions. The results generated for the control measures concept were evaluated against previous JHA documents, as the data were extracted from them, while the data generated for the remaining concepts were compared against the gold standard. The comprehensive evaluations demonstrated that the data migration process had effectively transferred the relevant instances to the O-JHAKG, while the JHAKG itself reliably and precisely retrieved the desired JHA information as per the input queries.

4. Discussion

It is clearly imperative that the JHA process should be automated to address the inherent limitations of the conventional process and to significantly improve the efficiency of the overall safety management process in the construction industry. As the JHA is often conducted as a brainstorming session, it relies heavily on the individuals' implicit knowledge derived from their extensive experience and expertise in the domain. Therefore, any approach focused on automating the JHA should not disregard the implicit knowledge possessed by these individuals. The proposed JHAKG was built considering both explicit and implicit knowledge in the JHA domain by conducting a JHA document review and a Delphi study to systematically extract that knowledge from the domain. This has resulted in a knowledge graph with an enhanced reasoning capacity as a semantic-based and rule-based inference mechanism was included, which was mostly derived by analysing transcripts of interviews with domain experts. To identify the hazards, the JHAKG considered the job step, the associated element and the execution method, while to identify the control measures, it took into account the potential consequence of the hazard, providing a more specific answer to the user. Moreover, to evaluate the risk of hazards, the JHAKG considered the external conditions present at the time to check the probability of the hazard's occurrence. Thus, the JHAKG has the capability to provide more specific information to the user, rather than providing generic information.

A rule-based NLP information extraction mechanism was adopted to populate the O-JHAKG with explicit information from incident reports. An initial screening process was employed to extract incidents that exhibited a semantic relationship between the resultant incident and the job step. Subsequently, a rule-based information extraction mechanism was implemented on the screened incident reports to extract the targeted information elements. This screening process significantly improved the performance of the information extraction mechanism and helped to achieve better precision, recall and F-measure values. Moreover, as information extraction heavily relies on syntactic features, the need to construct extensive keyword dictionaries was minimised.

5. Conclusion

Job hazard analysis (JHA) is considered the standard process for identifying hazards, thus playing a critical role in the safety management system. As a knowledge-intensive process, JHA heavily relies on the application and utilisation of knowledge. It typically involves analysing job steps and external conditions (such as human, environmental and management factors), employing critical thinking and leveraging accumulated knowledge to determine potential hazards, assess their risk levels and devise preventive measures. For these reasons, individuals performing JHA must rely on their experience, along with explicit knowledge sources such as previous JHAs, incident reports and Codes of Practice documents, within a limited time frame and, given the dynamic nature of the construction industry, they also must perform JHAs frequently.

Given these points, the current research proposed a knowledge-driven approach to perform JHA in a more specific manner by taking account of task-specific and environmental characteristics. A knowledge graph (i.e., the JHAKG) was developed to formalise the knowledge inherent in the JHA domain. The schema underlying the knowledge graph was built using METHONTOLOGY, a leading manual ontology development method. METHONTOLOGY facilitated the incorporation of implicit knowledge by enhancing the JHAKG's inferencing capacity. To enhance the comprehensiveness of the JHAKG, data from an incident database were integrated into the knowledge graph. A rule-based NLP approach was adopted to extract relevant information and instantiate the knowledge graph. The information extraction mechanism was validated using a gold standard. Performance evaluation metrics were calculated, achieving an F-measure for the concepts of over 0.90 which was considerably higher than that of other rule-based approaches. Finally, the instantiated knowledge graph was validated to test its comprehensiveness and inferencing capability. The results demonstrated that the developed JHAKG could assist safety personnel during the JHA process by providing quick and comprehensive safety risk information to effectively mitigate hazards.

This research study has contributed to both theory and practice. The conceptualisation step of the ontology development process has resulted in a detailed list of concepts and relationships that affect the comprehensive execution of the JHA process. This research has identified the attributes of entities that determine the existence of their relationships with other entities. Thus, it provides theoretical guidance to individuals who perform JHA, informing them about the variables they need to consider at each step to successfully complete the process. Furthermore, the resultant JHAKG can practically assist safety personnel during JHA by providing information based on previous JHAs and incident reports, enabling them to achieve quick and comprehensive outcomes. In terms of limitations, the proposed JHAKG can only evaluate the risk of primary hazards and does not provide risk evaluation for secondary hazards originating from external conditions. Furthermore, it only considers weather conditions, workplace conditions, atmospheric conditions and proximity conditions when evaluating the risk of primary hazards. Other external factors, such as human and management factors, could influence the risk of primary hazards; however, due to their subjectivity, they were excluded from the JHAKG reasoning mechanism.

Using semi-automated and automated methods seems to be a promising approach for developing knowledge-driven applications. Many researchers have utilised various machine learning techniques, such as unsupervised, supervised and deep learning techniques, to create knowledge graphs. With the ability to capture relationships between entities from large volumes of text, for future studies, machine learning techniques could be employed on databases, such as incident databases, to automatically generate knowledge graphs.

Acknowledgements

The authors would like to thank the South Australian Water Corporation (SA Water) and Mr John Persinos for the support given during this study.

Author Contributions

Sonali Pandithawatta contributed to the conceptualisation, methodology, software, validation, analysis, data collection, draft preparation and visualisation. Raufdeen Rameezdeen contributed to the conceptualisation, methodology, draft preparation, manuscript editing and supervision. Jun Ahn contributed to the conceptualisation, methodology, draft preparation, manuscript editing and supervision. Christopher W.K. Chow contributed to the draft preparation, manuscript editing and supervision and Nima Gorjian contributed to the draft preparation, manuscript editing and supervision.

Funding

This research received no specific financial support from any funding agency.

Institutional Review Board Statement

This project received approval from the University of South Australia's Human Research Ethics Committee as documented in Ethics Protocol 204341.

References

Albrechtsen, E., Solberg, I., and Svensli, E. (2019). The application and benefits of job safety analysis. *Safety Science,* 113, 425–437. doi:10.1016/j.ssci.2018.12.007

Altawil, S. N. A. (2017). *Multi-Level Knowledge Extraction and Modeling to Support Job Hazard Analysis Process for Oil and Gas Pipeline Projects.* (Doctor of Philosophy in Construction Engineering and Management). University of Alberta.

Asadzadeh, A., Arashpour, M., Lib, H., Ngoc, T., Bab-Hadiashard, A., and Rashidie, A. (2020). Sensor-based safety management. *Automation in construction,* 113. doi:10.1016/j.autcon.2020.103128

Brusa, G., Caliusco, M. L., and Chiotti, O. (2006). *A process for building a domain ontology: An experience in developing a government budgetary ontology*. Paper presented at the Conferences in Research and Practice in Information Technology, Hobart, Australia.

Chen, J., Lu, W., Fu, Y., and Dong, Z. (2023). Automated facility inspection using robotics and BIM: A knowledge-driven approach. *Advanced Engineering Informatics,* 55. doi:10.1016/j.aei.2022.101838

Chi, N. W., Lin, K. Y., and Hsieh, S. H. (2014). Using ontology-based text classification to assist Job Hazard Analysis. *Advanced Engineering Informatics,* 28(4), 381-394. doi:10.1016/j.aei.2014.05.001

Cooke, T., Lingard, H., Blismas, N., and Stranieri, A. (2008). ToolSHeDTM: The development and evaluation of a decision support tool for health and safety in construction design. *Engineering, Construction and Architectural Management,* 15(4), 336-351. doi:10.1108/09699980810886847

Crnkovic, G. D. (2010). Constructive Research and Info-computational Knowledge Generation. In *Model-Based Reasoning in Science and Technology*, Berlin: Springer.

Crutchfield, N., and Roughton, J. (2019). Using the Job Hazard Analysis. In *Safety Culture*, Elsevier, 289-325.

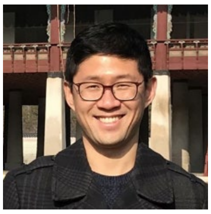
D'Avanzo, E., Lieto, A., and Kufik, T. (2008). *Manually vs semiautomatic domain specifc ontology building.* (Thesis in Information and Commercial Electronics). University of Salerno, Retrieved from https://www.academia.edu/25643163/Manually_vs_semiautomatic_domain_specific_ontology_building
Dayé, C. (2018). How to train your oracle: The Delphi method and its turbulent youth in operations research and the policy sciences. *Social Studies of Science,* 48(6), 846–868. doi:10.1177/0306312718798497
Ding, L. Y., Zhong, B. T., Wu, S., and Luo, H. B. (2016). Construction risk knowledge management in BIM using ontology and semantic web technology. *Safety Science,* 87, 202-213. doi:10.1016/j.ssci.2016.04.008
Elmhadhbi, L., Mohamed-Hedi, K., Archimède, B., Otte, J. N., and Smith, B. (2021). An Ontological Approach to Enhancing Information Sharing in Disaster Response. *Information,* 12(10). doi:10.3390/info12100432
Fernández-López, M., Gómez-Pérez, A., and Juristo, N. (1997). *METHONTOLOGY: From ontological art towards ontological engineering*. Paper presented at the Fourteenth national conference on artificial intelligence.
Gasevic, D., Djuric, D., and Devedzic, V. (2009). *Model Driven Engineering and Ontology Development*: Springer Verlag Berlin Heidelberg.
Glenn, D. D. (2011). Job safety analysis. *Professional Safety,* 56(4), 48-57.
Goh, Y. M., and Chua, D. K. H. (2009). Case-based reasoning for construction hazard identification: Case representation and retrieval. *Journal of construction engineering and management,* 135(11), 1181-1189. doi:10.1061/(ASCE)CO.1943-7862.0000093
Goh, Y. M., and Chua, D. K. H. (2010). Case-based reasoning approach to construction safety hazard identification: Adaptation and utilization. *Journal of Construction Engineering and Management,* 136(2), 170-178. doi:10.1061/(ASCE)CO.1943-7862.0000116
Harteis, C. (2022). Delphi-Technique as a Method for Research on Professional Learning. In *Methods for Researching Professional Learning and Development*, 33, 351–371.
Hillege, S. L., van Domburgh, L., Mulder, E. A., Jansen, L. M. C., and R.R.J.M., V. (2018). How do forensic clinicians decide? A Delphi approach to identify domains commonly used in forensic juvenile treatment planning. *International Journal of Offender Therapy and Comparative Criminology,* 62(3), 591–608. doi:10.1177/0306624X16658724
Kasanen, E., Lukka, K., and Siitonen, A. (1993). The constructive approach in management accounting research. *Journal of Management Accounting Research,* 5, 241-264.
Khatoon, A., Hafeez, Y., and Ali, T. (2014). An ontological framework for requirement change management in distributed environment. *Nucleus,* 46(3), 291-301.
Lehtiranta, L., Junnonen, J.-M., Kärnä, S., and Pekuri, L. (2015). Research Approach: Problem Solving for Complex Projects. In B. Pasian and R. Turner (Eds.), *Designs, Methods and Practices for Research of Project Management*. London: Taylor and Francis
Lukka, K. (2000). The key issues of applying the constructive approach to field research. In T. Reponen (Ed.), *Management Expertise for the New Millennium*: Turku School of Economics and Business Administration.
McGregor, C. (2018). Using constructive research to structure the path to transdisciplinary innovation and its application for precision public health with big data analytics. *Technology Innovation Management Review,* 8(8), 7-15. doi:10.22215/timreview/1174
Mitropoulos, P., Abdelhamid, T. S., and Howell, G. A. (2005). Systems Model of Construction Accident Causation. *Journal of construction engineering and management,* 131(7).
Moghadami, M., and Mortazavi, A. (2018). Development of a Risk-based Methodology for Rock Slope Analysis. *International Journal of Civil Engineering,* 16. doi:10.1007/s40999-017-0246-z
Naisola-Ruiter, V. (2022). The Delphi technique: a tutorial. *Research in Hospitality Management,* 12(1), 91-97. doi:10.1080/22243534.2022.2080942
Niederberger, M., and Köberich, S. (2021). Coming to consensus: the Delphi technique. *European Journal of Cardiovascular Nursing,* 20(7), 692-695. doi:10.1093/eurjcn/zvab059
Pandithawatta, S., Ahn, S., Rameezdeen, R., Chow, C. W. K., Gorjian, N., and Kim, T. W. (2023). Development of a Knowledge Graph for Automatic Job Hazard Analysis: The Schema. *Sensors,* 23(8). doi:10.3390/s23083893
Poghosyan, A., Manu, P., Mahamadu, A. M., Akinade, O., Mahdjoubi, L., Gibb, A., and Behm, M. (2020). A web-based design for occupational safety and health capability maturity indicator. *Safety Science,* 122. doi:10.1016/j.ssci.2019.104516
Rosson, M. B., and Carroll, J. M. (2002). *Usability engineering: scenario-based development of human-computer interaction*. San Francisco, California: Morgan Kaufmann.
Spickermann, A., Zimmermann, M., and von der Gracht, H. A. (2014). Surface- and deep-level diversity in panel selection — Exploring diversity effects on response behaviour in foresight. *Technological Forecasting and Social Change,* 85, 105-120. doi:10.1016/j.techfore.2013.04.009
Sternberg, R. J., Forsythe, G. B., Hedlund, J., Horvath, J. A., Wagner, R. K., Williams, W. M., . . . Grigorenko, E. L. (2000). *Practical Intelligence in Everyday Life*. New York, NY, US: Cambridge University Press.
Thompson, J. (2022). A Guide to Abductive Thematic Analysis. *The Qualitative Report,* 27(5). doi:10.46743/2160-3715/2022.5340
Vaticle. (2023). TypeDB.
Wang, H.-H., and Boukamp, F. (2011). Ontology-based representation and reasoning framework for supporting job hazard analysis. *Journal Of Computing In Civil Engineering,* 25(6), 442-456. doi:10.1061/(ASCE)CP.1943-5487.0000125

Xing, X., Zhong, B., Luo, H., Li, H., and Wu, H. (2019). Ontology for safety risk identification in metro construction. *Computers in Industry,* 109, 14-30. doi:10.1016/j.compind.2019.04.001

Zhang, S., Boukamp, F., and Teizer, J. (2015). Ontology-based semantic modeling of construction safety knowledge: Towards automated safety planning for job hazard analysis (JHA). *Automation in Construction,* 52, 29-41. doi:10.1016/j.autcon.2015.02.005

Sonali Pandithawatta is a final-year PhD student at the University of South Australia, focusing her research on developing an automated solution to assist in job hazard analysis for construction operations using a knowledge-driven approach. She earned her undergraduate degree in Quantity Surveying from the University of Moratuwa, Sri Lanka. Her research interests include Knowledge Management and Representation, Construction Health and Safety, Natural Language Processing Applications in Construction Information Management, Green Construction, and Lean Construction. She is currently employed as a casual lecturer at the University of South Australia.

Raufdeen Rameezdeen is a Professor of Construction and Project Management at UniSA STEM. He holds a PhD in Infrastructure Systems Engineering, and his research explores the intersections of technological solutions, economic efficiency, and environmental sustainability in the development and management of infrastructure.



Seungjun Ahn is an Assistant Professor at Hongik University. He obtained his PhD from the University of Michigan, USA, and his research interests include construction safety, knowledge management and representation, NLP applications in construction information management, and various fields related to construction project management.

Christopher W.K. Chow is a Professor of Water Science and Engineering at UniSA STEM. He obtained his PhD in Analytical Chemistry from the University of South Australia. His research interests include real-time optimization of water treatment processes, disinfection control, disinfection by-products, and distribution system management and modeling.

Nima Gorjian is an Adjunct Associate Research Professor at UniSA STEM. He obtained his PhD from Queensland University of Technology, Australia, and his research interests lie in water engineering, construction safety, and various fields relevant to construction project management.