

From Worker Narratives to Safety Intelligence: Extracting Construction Health Risks Using NLP and Knowledge Graphs

Karen DSouza¹, Yuting Chen², Gongfan Chen², Lingguang Song² and Don Chen²

¹School of Science, Department of Computer Science, Purdue University, U.S.A.

²Albert School of Construction, Department of Engineering Technology and Construction Management, University of North Carolina at Charlotte, U.S.A.

kmdsouza@purdue.edu, ychen106@charlotte.edu,
gchen12@charlotte.edu, lsong2@charlotte.edu, dchen9@charlotte.edu

Abstract -

Construction workers face multifaceted physical and psychological health risks due to challenging and often unpredictable working environments. This study introduces a computational framework that uses artificial intelligence and natural language processing to extract and structure occupational health insights from 7,441 Reddit posts and comments related to construction. Focusing on eight custom entity categories that include worker types, health symptoms, workplace stressors, institutional actions, and control measures, we applied rule-based entity recognition methods to identify recurring patterns and concerns discussed by workers. Frequency analysis highlighted prevalent health symptoms such as stress, dehydration, and exhaustion. A domain-specific knowledge graph for stressors, health symptoms and control measures revealed relationships between extracted entities, uncovering structural patterns, isolated risks, and frequently co-mentioned hazards. The framework demonstrates how informal, user-generated narratives can be transformed into structured knowledge to support occupational health surveillance. This work contributes scalable tools that can assist policymakers, unions, safety officers, and researchers in understanding evolving worker needs. By bridging Natural Language Processing (NLP) techniques with occupational safety, our method offers a new lens for data-driven intervention strategies rooted in real-world discourse.

Keywords -

Knowledge Graphs; Web Scraping, Natural Language Processing; Physical Health; Construction Workers; Risk Factors

1 Introduction

The construction industry is consistently ranked among the most hazardous occupational sectors worldwide, with workers frequently exposed to a combination of physical, environmental, and psychological stressors [1]. Despite extensive safety regulations and awareness campaigns, the dynamic and unpredictable nature of construction sites continues to pose serious health risks [2]. Traditional ap-

proaches to occupational health and safety rely heavily on formal incident reports, regulatory citations, and clinical diagnoses [3][4]. These reports often fail to capture the lived experiences of frontline workers. The sources may underreport certain types of safety issues such as physical health conditions, mental health struggles, and risk factors that lead to gaps in intervention and prevention strategies [5].

At the same time, the social digital platforms such as Reddit and X have introduced new avenues for workers to share uncensored, real-time narratives about their occupational experiences. These narratives include valuable information about risk exposure, institutional responses, and perceived gaps in workplace protections [6]. However, due to their informal and unstructured nature, such data sources are largely untapped in occupational safety research [7]. The unstructured and noisy nature of this data presents significant challenges for systematic analysis and interpretation.

Existing work in occupational health surveillance using natural language processing (NLP) has demonstrated the potential of social media to supplement official data [8]. Yet, few studies have tailored these techniques specifically to the construction context, and even fewer have attempted a structured, multi-entity analysis using domain-specific categories relevant to physical and mental health, job roles, and safety interventions. This study addresses these gaps by developing a computational framework for extracting and analyzing occupational risk narratives using NLP and named entity recognition (NER).

In recent years, NLP and knowledge representation techniques have shown promising potential in structuring large-scale unstructured text data and generative artificial intelligence [9]. Among these, knowledge graphs which as graph-based representations that encode semantic relationships among entities have emerged as powerful tools for extracting and analyzing contextual patterns from narrative data [10][11]. Coupled with transformer-based models for NER and relation extraction, knowledge graphs enable analyzable structures that reflect the complexity of

occupational discourse [12].

In this study, we propose a computational framework that leverages NLP and knowledge graph construction to analyze 7,441 Reddit posts and comments related to the construction sector. We first define a construction domain-specific schema encompassing eight key entities such as worker types, environmental stressors, health symptoms, and institutional actions. We then use these categories to extract and map relationships within the text. Our approach enables the discovery of emerging risk clusters, reveals central actors and symptoms within the discourse. The corresponding knowledge graphs highlight underreported hazards such as heat related health symptoms, and inadequate hydration practices.

2 Related Work

Occupational safety in construction has been studied extensively through traditional sources such as injury reports, surveys, and regulatory databases [13]. However, these methods often suffer from underreporting and lack of real-time insight into worker experiences. Recent research has started to explore alternative data sources, including social media platforms and online forums, as a means of capturing informal yet rich narratives about workplace conditions.

2.1 Social Media and Informal Occupational Data

The use of social media as a data source for occupational health surveillance is a growing field. Recent work done by researchers has (i) explored mining Reddit posts for mental health indicators, highlighting the platform’s potential for early detection of stress and burnout. (ii) researchers applied NLP techniques to datasets from social media such as X data to monitor workplace hazards. These studies reinforce the value of informal, crowd-sourced narratives in complementing official safety reporting mechanisms.

For risk analysis, work has been done in the construction industry with researchers evaluating failure modes and effects analysis as well as highlighting the overall risks analysis [1][14][15][16].

2.2 Knowledge Graphs and Occupational Safety

Knowledge graphs have gained popularity as a structured approach for modeling relationships between concepts in unstructured text. In occupational safety contexts, these graphs have been used to map hazard domains, link injury types to causes, and integrate data from disparate sources for improved decision-making [17][18].

Domain adaptation of NER techniques has become crucial in contexts like construction, where entities may not align with those in general purpose domains. Hybrid approaches combining dictionary based methods and trans-

former models like BERT have proven effective in extracting occupationally relevant entities such as symptoms, worker roles, and safety measures. Beyond entity identification, relation extraction enables the construction of semantic graphs that reveal causal and correlational patterns.

3 Methodology

The data collection and processing pipeline framework is a crucial part of the project. The quality of the results depends heavily on the data processing techniques used.

3.1 Data Collection and Processing

To analyze occupational risk narratives from the construction sector, we extracted Reddit posts and comments using a keyword-based scraping strategy. Keywords were selected to reflect both domain-specific roles (e.g. construction, worker, airport) and health-related stressors (e.g. burnout, depression, heat, noise, fatigue, mental). These combinations were used to query Reddit via the PRAW API, resulting in over 130,000 subreddit matches. To increase domain relevance, filtering was applied to retain only posts from subreddits such as Construction, ConstructionManagers, HVAC, Plumbing, and civilengineering, yielding a dataset of 7,441 posts and comments.

All retrieved text was normalized to lowercase for consistency, and a manual filtering step was conducted to remove irrelevant matches, such as metaphorical uses of construction or posts unrelated to workplace safety. The dataset was filtered to retain only English-language text and exclude irrelevant or low-information content such as memes, spam, or non-descriptive titles. All text entries were cleaned to remove links, special characters, and excessive whitespace. Sentences were tokenized and split using standard NLP preprocessing techniques. This preparation was essential to ensure high-quality input for entity recognition and relation extraction in the subsequent steps.

3.2 Annotation

The next step was the data annotation process. A structured annotation protocol was applied to a corpus of 7,441 Reddit posts and comments related to construction work. The dataset consisted of worker types that included categories such as construction, postal, delivery, industrial and so on. However, the annotation focused on a key dimension: relevance to construction worker type data only. The process involved three annotators with domain knowledge to ensure robustness, minimize subjectivity, and support cross-validation.

The cross-validation annotator step was included as the last step in the annotation process. The cross-validation

notes included justifications for labeling decisions, highlighting ambiguity, context-specific interpretations, and suggestions for edge-case resolution. For instance, in one entry, physical strain was indicated by indirect language, prompting a note about interpretation boundaries.

The manual annotation serves as a ground truth foundation for subsequent entity extraction and relation modeling. It allows the computational models to be evaluated or trained with human verified categories and enables higher-quality graph construction by filtering out noise and irrelevance from informal social media narratives.

3.3 Named Entity Recognition (NER)

To extract structured knowledge from unstructured text data related to construction worker health and safety, we implemented a custom NER pipeline. Our approach relied on rule-based pattern matching, lemmatization, and curated dictionaries mapped to eight distinct entity categories relevant to occupational safety, health, and organizational behavior. We based our work on a subset of 525 narratives which were cross-validated targeting construction-specific content extracted from a larger corpus of 7,441 worker incident narratives.

Domain-specific dictionaries and spaCy's Phrase-Matcher were used to supplement transformer-based models, allowing for more accurate recognition of low-frequency or colloquial phrases that may not appear in standard corpora. All entities were normalized to lower-case to reduce duplication and unify mentions. Next, we describe the methodology and criteria used for each of the eight entity categories. These categories are listed Table 1 along with representative examples.

Table 1. Entity Categories

Category	Examples
Worker Type Managers	Supervisor, project manager
Worker Type Frontline	Electrician, roofer
Health Symptoms	Fatigue, dizziness
Environmental Stressors	Heat, cold
Workplace Stressors	Long shift, understaffing
Individual Factors	Age, single
Institutional Actions	OSHA inspection, fine
Control Measures	Safety training
Additional Category	Example
Location	Dallas, Houston

3.3.1 Worker Types

The worker type category was divided into two subtypes: (i) frontline workers and (ii) managers. Frontline worker entities such as electrician, roofer, plumber, and

laborer represent skilled tradespeople directly involved in physical construction tasks. They constituted **worker type frontline**. Managerial worker entities such as supervisor and project manager reflect administrative roles within construction settings. They were included in **worker type managers**. To improve precision and reduce ambiguity, these entity mentions were matched and normalized using occupation classification terms and examples referenced from the U.S. Bureau of Labor Statistics (BLS.gov) [19]. This alignment with standardized occupational definitions ensured that extracted roles mapped consistently to recognized construction labor segments.

3.3.2 Health Symptoms

Health symptoms were extracted using a dictionary-based lemmatized matcher for both physical and psychological health indicators. Common entities included fatigue, exhaustion, headache, dehydration, stress, and burnout. The matched symptom entities were later used in co-occurrence and knowledge graph-based analyses.

3.3.3 Environmental Stressors

Environmental stressors refer to physical conditions present on the job site. We built a controlled vocabulary around terms such as noise, heat, sunlight, chemical, dust, and air quality. Each term was identified using lemmatized rule-based matching, often in combination with contextual qualifiers. For example, sun exposure and UV radiation were collapsed under sunlight. These stressors were crucial in identifying the relationship between job site conditions and adverse health outcomes.

3.3.4 Workplace Stressors

This category focused on psychosocial and organizational stressors. Examples include harassment, trauma, overwork, pressure, and lack of support. Unlike environmental stressors, these were often mentioned in indirect language or idiomatic phrases (e.g. toxic work culture). We expanded the dictionary with variations and synonyms using pre-trained word embeddings to improve recall. Workplace stressors were shown to have strong co-occurrence with psychological symptoms such as anxiety and burnout.

3.3.5 Individual Factors

Personal vulnerabilities such as pre-existing conditions or life circumstances were grouped under this category. Detected terms included older worker, health issues, disability, and family responsibilities. These entities help provide a nuanced understanding of differential risk exposure.

3.3.6 Institutional Actions

This category captured formal institutional responses such as OSHA investigation, union involvement, citations, lawsuits, and fines. A regular expression and phrase-based matching system was used to detect regulatory terms and legal follow-ups. These entities were often connected to either severe outcomes or recurring risk scenarios and served as anchor nodes in our knowledge graph to highlight systemic issues.

3.3.7 Control Measures

We extracted procedural mitigation measures such as hydration, shade, rest breaks, safety training, and cooling measures. This category showed how actively safety protocols were implemented post-incident. Entity detection relied on a mix of procedural keywords and compliance terms sourced from OSHA and construction industry best practices.

Risk mitigation strategies were divided into 5 subcategories called elimination, substitution, engineering controls, administrative controls, personal protective equipment (PPE). Entities were tagged accordingly and used to evaluate the adequacy of mitigation strategies in each case. This tiered tagging allowed assessment of how often high-level (elimination, substitution) versus low-level PPE controls were employed.

3.3.8 Additional Categories

We identified one additional category (i) location. To preprocess location entities, geographic references within the text data were extracted using NER with spaCy's pre-trained language model. The model detects entities tagged as GPE (Geo-Political Entities), which include names of cities, states, and countries. Each row of the dataset was processed to identify these locations. After extraction, the results were aggregated and filtered to compile a distinct list of frequently mentioned locations. This method allowed for both broad and granular recognition of geographic context capturing entries like Texas, Dallas and Travis County. The compiled list of location entities plays a key role in constructing a spatial layer within the knowledge graph, enabling location-aware analysis of occupational safety concerns across different regions.

Each recognized entity in the 8 categories and the 2 additional categories were stored as a list in a new column later converted into binary label vectors (1 if the entity was present in a document, 0 otherwise) for downstream modeling and correlation analysis.

3.4 Knowledge Graphs

The motivation to use knowledge graphs stems from the need to model complex, non-linear relationships in unstructured text. Reddit narratives often describe multifactorial risks, where a single worker may report a combination of symptoms caused by several stressors, within a particular environmental or organizational context. Traditional text classification or keyword-based techniques cannot represent these many-to-many, context-dependent interactions. In this context, knowledge graphs offer the following advantages in the construction domain:

- **Semantic structuring:** Graphs preserve the relational meaning between entities, helping to reveal patterns such as co-occurrence of symptoms and institutional inaction.
- **Interpretability:** Unlike black-box models, graphs allow human analysts and stakeholders to visually trace paths between causes and effects.
- **Graph analytics:** We can apply network algorithms such as centrality, clustering, and shortest-path analysis to identify influential symptoms or recurring stressor combinations.
- **Scalability:** As new data is collected, it can be continuously incorporated into the graph without retraining, enabling real-time risk tracking.

3.5 Rule-Based Symptom Extraction Performance Evaluation

To evaluate the performance of the rule-based entity extraction system, a subset of annotated data was used where relevance was confirmed by a primary annotator, resulting in 525 records. For the symptom category, binary ground truth labels were manually verified and compared against predictions generated by the rule-based matching approach. The presence of symptom-related entities was treated as a document-level classification task, where each record was labeled as either containing a symptom (1) or not (0). Precision, recall, and F1-score were computed to assess performance.

To compare rule-based extraction with a machine learning baseline, a BERT-based classifier was fine-tuned for binary health symptom extraction using the same annotated dataset. As shown in Table 3, the rule-based approach achieved perfect recall (1.00), indicating that all symptom-related instances were successfully identified. However, it exhibited slightly lower precision due to false positives. In contrast, the BERT model achieved a more balanced performance, with an F1-score of 0.91, by leveraging contextual understanding to reduce over-detection. While BERT improved precision-recall balance, the rule-based method remains advantageous for high-recall applications and interpretability. To ensure a fair comparison, both the rule-based approach and the BERT model were evaluated on the same held-out test set.

Table 2. Performance of Rule-Based Health Symptom Extraction

Class	Precision	Recall	F1-score
No Health Symptoms (0)	1.00	0.63	0.77
Health Symptoms (1)	0.89	1.00	0.94
Accuracy		0.91	
Macro Avg	0.95	0.81	0.86
Weighted Avg	0.92	0.91	0.91

Table 3. Comparison of Rule-Based and BERT Models for Health Symptoms

Model	Precision	Recall	F1-score
Rule-Based	0.89	1.00	0.94
BERT	0.87	0.96	0.91

4 Findings and Analysis

4.1 Frequency Analysis of Extracted Entities

After completing entity extraction, we analyzed the frequency of each recognized term within its respective category to identify dominant concerns and patterns in the construction domain.

For each entity column, the lists of matched keywords were flattened and aggregated to compute their frequency. We plotted the top 10 most common terms for each of the eight categories. This provided insight into which symptoms, stressors, and interventions were most frequently mentioned in the dataset. The findings from frequency analysis are described below. These frequency distributions served as the foundation for more advanced relational analysis through co-occurrence heatmaps and graph modeling. Figure 1 shows keywords like laborer, plumber, roofer and electrician among the top entities in frontline worker type. The environmental stressors in Figure 2 shows top entities like heat, cold, noise and sunlight. The top healthcare symptoms in Figure 3 shows stress, dehydration, exhaustion, and physical labor.

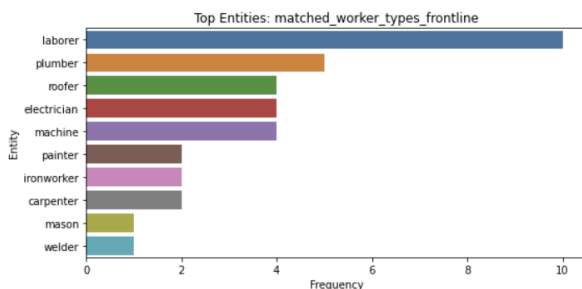


Figure 1. Frequency of Worker type frontline

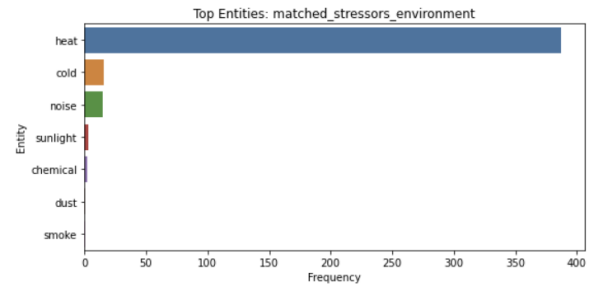


Figure 2. Frequency of Environment Stressors

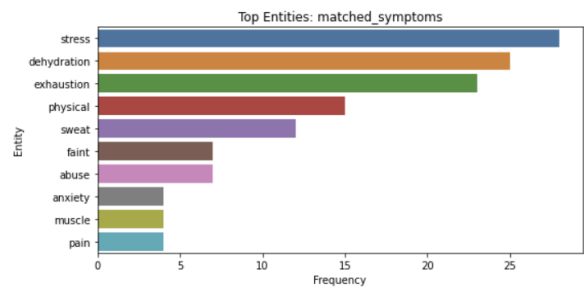


Figure 3. Frequency of Health Symptoms

4.2 Co-occurrence and Correlation Analysis

We next analyzed how frequently entities from different categories appeared together within the same document using a co-occurrence matrix. For example, we examined whether certain symptoms (e.g. burnout) were consistently mentioned alongside specific stressors (e.g. overwork or lack of support). The analysis computed pairwise co-occurrence counts between entity labels using binary-encoded columns. The resulting matrix was visualized as a heatmap, where color intensity indicates higher co-occurrence frequency.

- Heat showed the highest co-occurrence with work, reinforcing environmental causality.
- Heat correlated heavily with job, hourly.
- Heat was correlated highly with stress, injury, health.

This correlation structure informed the next step of constructing a knowledge graph to visualize entity relationships.

4.3 Insights from the Knowledge Graph

To better understand how extracted entities interact, we constructed an integrated knowledge graph linking stressors, health symptoms, and control measures. These three categories were selected to represent a conceptual pathway

from exposure to outcome to response. The knowledge graph in Figure 4 captures the co-occurrence of stressors, health symptoms and control measures in Reddit narratives, with node size indicating entity frequency and edge thickness reflecting co-mention strength. This visualization uncovers significant patterns. In this framework, stressors (e.g., heat, cold) represent environmental or organizational risk factors, health symptoms capture the resulting physical or psychological effects, and control measures represent mitigation strategies discussed in worker narratives.

By connecting these categories within a single graph, Figure 4 enables the identification of complete risk-to-response pathways rather than isolated associations. This provides a more holistic view of occupational health dynamics, illustrating how specific stressors lead to particular health outcomes and how those outcomes are addressed through practical interventions.

The integrated knowledge graph reveals several dominant stressor–symptom–control pathways.

- The most prominent pathway is heat → heat exhaustion → rest breaks, indicating that acute thermal stress is frequently associated with rest-based mitigation strategies. Rest breaks are a control measure.
- Similarly, heat → dehydration → hydration highlights the importance of water access as a primary response to heat exposure. Hydration is a control measure.
- Another recurring pathway, heat → physical exhaustion → rest breaks, suggests that prolonged exposure leads to fatigue, which is addressed through rest-based interventions.
- In addition to heat-dominated pathways, other environmental relationships are also observed. For example, the pathway sunlight → fainting → shade highlights how direct solar exposure can lead to acute physiological responses that are mitigated through access to shaded rest areas. This reinforces the importance of environmental controls alongside behavioral interventions in managing occupational heat stress.

These pathways demonstrate that worker-reported experiences follow consistent patterns linking environmental stressors to physiological responses and subsequent mitigation actions. The integrated knowledge graph reveals incidents of heat-related health risk severity reflected in worker narratives. Early indicators of thermal stress include symptoms such as sweating and fatigue, which are associated with initial mitigation measures such as cooling and rest. Symptoms such as dehydration emerge, often linked to hydration-based interventions. More severe

conditions, including vomiting and physical exhaustion, indicate escalating physiological stress. In extreme cases, critical symptoms such as fainting are observed, typically associated with immediate protective measures such as access to shaded areas.

This progression highlights how occupational heat exposure manifests across multiple levels of severity and indicates the control measures that were reported in Reddit narratives.

4.4 Limitations

Reddit provides access to rich, real-time worker narratives, it has several limitations as a data source. The platform’s user base may not be representative of the broader construction workforce, as it is likely skewed toward younger, more digitally engaged individuals. Additionally, the content is self-reported and may reflect subjective perceptions rather than verified incidents. As a result, the findings should be interpreted as indicative patterns of worker experiences rather than statistically representative conclusions. Despite these limitations, Reddit offers valuable insights into underreported issues such as fatigue, burnout, and psychosocial stress, complementing traditional safety reporting systems.

5 Conclusion

This work demonstrates how unstructured social media narratives can be transformed into structured occupational safety insights using a computational framework grounded in NER and knowledge graph modeling. By extracting eight domain-specific entity categories related to worker roles, symptoms, stressors, institutional actions, and mitigation strategies, we reveal risk patterns such as heat-related symptoms, fatigue, and psychological stress in construction settings. Frequency and co-occurrence analysis highlight strong relationships between environmental and workplace stressors and adverse health outcomes. The resulting knowledge graph exposes systemic gaps between health symptoms and workplace stressors. Together, these findings illustrate the value of integrating informal worker discourse into scalable occupational health surveillance and lay the groundwork for future transformer-based models to enhance precision, generalizability, and policy relevance.

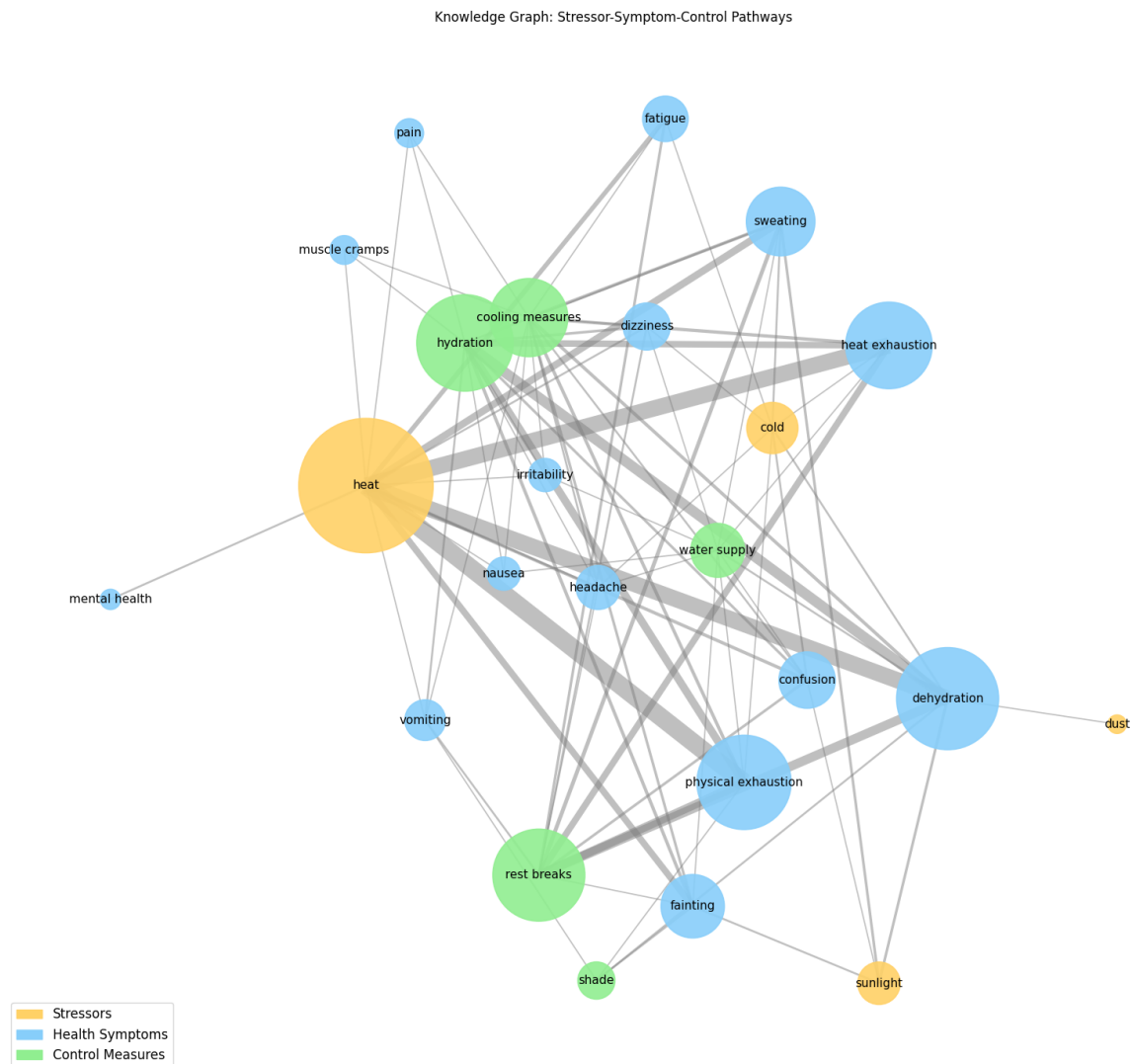


Figure 4. Knowledge Graph: Stressor-Symptoms-Control Pathways. Knowledge Graph illustrates integrated pathways between stressors (yellow), health symptoms (blue), and control measures (green). Nodes represent extracted entities, and edges indicate co-occurrence within worker narratives. The graph highlights dominant pathways such as heat → dehydration → hydration and heat → physical exhaustion → rest breaks, demonstrating how occupational risks are experienced and mitigated in practice.

References

- [1] Devdatt P Purohit, NA Siddiqui, Abhishek Nandan, and Bikarama P Yadav. Hazard identification and risk assessment in construction industry. *International Journal of Applied Engineering Research*, 13(10): 7639–7667, 2018.
- [2] KA Shamsuddin, Mohd Norzaimi Che Ani, AK Ismail, and MR Ibrahim. Investigation the safety, health and environment (she) protection in construction area. *International Research Journal of Engineering and Technology*, 2(6):624–636, 2015.
- [3] Johan M Sanne. Incident reporting or storytelling? competing schemes in a safety-critical and hazardous work setting. *Safety Science*, 46(8):1205–1222, 2008.
- [4] Heather M Zoller. Health on the line: Identity and disciplinary control in employee occupational health and safety discourse. *Journal of Applied Communication Research*, 31(2):118–139, 2003.
- [5] Nadia A Al-Aubaidy, Carlos H Caldas, and Stephen P Mulva. Assessment of underreporting factors on construction safety incidents in us construction projects.

- International journal of construction management*, 22(1):103–120, 2022.
- [6] Munmun De Choudhury and Scott Counts. Understanding affect in the workplace via social media. In *Proceedings of the 2013 conference on Computer supported cooperative work*, pages 303–316, 2013.
 - [7] JY Teoh. How occupational therapy practitioners use virtual communities on the facebook social media platform for professional learning: A critical incident study. *Scandinavian Journal of Occupational Therapy*, 29(1):58–68, 2022.
 - [8] Michael Paul and Mark Dredze. You are what you tweet: Analyzing twitter for public health. In *Proceedings of the international AAAI conference on web and social media*, volume 5, pages 265–272, 2011.
 - [9] Alexis Figueroa-Baltazar and Anh D Chau. Leveraging generative ai to analyze construction industry trends through social media data. In *Computing in Civil Engineering 2025: Computational and Intelligent Technologies*, pages 36–45.
 - [10] Aidan Hogan, Eva Blomqvist, Michael Cochez, Claudia d’Amato, Gerard De Melo, Claudio Gutierrez, Sabrina Kirrane, José Emilio Labra Gayo, Roberto Navigli, Sebastian Neumaier, et al. Knowledge graphs. *ACM Computing Surveys (Csur)*, 54(4):1–37, 2021.
 - [11] Shaoxiong Ji, Shirui Pan, Erik Cambria, Pekka Marttinen, and Philip S Yu. A survey on knowledge graphs: Representation, acquisition, and applications. *IEEE transactions on neural networks and learning systems*, 33(2):494–514, 2021.
 - [12] Tareq Al-Moslmi, Marc Gallofré Ocaña, Andreas L Opdahl, and Csaba Veres. Named entity extraction for knowledge graphs: A literature overview. *IEEE access*, 8:32862–32881, 2020.
 - [13] Zhipeng Zhou, Yang Miang Goh, and Qiming Li. Overview and analysis of safety management studies in the construction industry. *Safety science*, 72:337–350, 2015.
 - [14] Hao-Tien Liu and Yieh-lin Tsai. A fuzzy risk assessment approach for occupational hazards in the construction industry. *Safety science*, 50(4):1067–1078, 2012.
 - [15] Abel Pinto, Isabel L Nunes, and Rita A Ribeiro. Occupational risk assessment in construction industry—overview and reflection. *Safety science*, 49(5):616–624, 2011.
 - [16] Leslie Edwards and Leslie J Edwards. *Practical risk management in the construction industry*. Thomas Telford, 1995.
 - [17] Qi Yao, Rita Yi Man Li, Lingxi Song, and M James C Crabbe. Construction safety knowledge sharing on twitter: A social network analysis. *Safety science*, 143:105411, 2021.
 - [18] Wenjing Wu, Caifeng Wen, Qi Yuan, Qiulan Chen, and Yunzhong Cao. Construction and application of knowledge graph for construction accidents based on deep learning. *Engineering, construction and architectural management*, 32(2):1097–1121, 2025.
 - [19] U.S. Bureau of Labor Statistics. Standard occupational classification (soc) system, 2018. URL <https://www.bls.gov/soc/2018/home.htm>. Accessed: 2026-01-31.