Identifying the emerging vulnerability of railway transport systems across countries by automated analysis of railway accident reports

Wei-Ting Hong

A thesis submitted in fulfillment of the requirements for the degree of

Doctor of Philosophy

Institute of Transport and Logistics Studies

University of Sydney Business School

The University of Sydney

November 2023

Dedicated to my wife and daughter, Ching-Ya (Bella) and Hsin-Yueh (Luna),
I won't come this far without your support.

Author attribution

The work delivered in the body of this thesis, except otherwise acknowledged, is the result of my own research investigations.

Chapter 2 contains materials used in the paper:

Hong, W. T., Clifton, G., & Nelson, J. D. (2022). Rail transport system vulnerability analysis and policy implementation: Past progress and future directions. ***Transport Policy***, 128, 299–308. https://doi.org/https://doi.org/10.1016/j.tranpol.2022.02.004

Chapter 3 contains materials used in the paper:

Hong, W. T., Clifton, G., & Nelson, J. D. (2023). Railway accident causation analysis: Current approaches, challenges and potential solutions. ***Accident Analysis & Prevention***, 186, 107049. https://doi.org/https://doi.org/10.1016/j.aap.2023.107049

Chapter 4, Chapter 5 and Chapter 6 contain material that has been integrated into two individual papers:

Hong, W. T., Clifton, G., & Nelson, J. D. (2023). RecoMap - a semi-automated tool for analysing railway accident recommendations across jurisdictions and over time. Submitted to ***Accident Analysis & Prevention***

Hong, W. T., Clifton, G., & Nelson, J. D. (2023). A data-driven conceptual framework for understanding the nature of hazards in railway accidents. Submitted to ***Transport Policy***

Chapter 7 contains materials used in the paper:

Hong, W. T., Clifton, G., & Nelson, J. D. (2023). Railway safety deterioration: an underlying potential hazard. Submitted to ***Accident Analysis & Prevention***

Part of this thesis was presented at the following conferences:

Hong, W. T., Clifton, G., & Nelson, J. D. (2023). A conceptual framework for the automated analysis of railway accident reports. Australasian Transport Research Forum 2023 Proceedings.

Hong, W. T., Clifton, G., & Nelson, J. D. (2023). Railway safety deterioration as a potential hazard. Transport Research Association for New South Wales Symposium 2023.

Hong, W. T., Clifton, G., & Nelson, J. D. (2022). Utilising natural language processing to learn from history and disseminate railway safety knowledge. Transport Research Association for New South Wales Symposium 2022. Retrieved from https://tinyurl.com/3u5njdsu

Hong, W. T., Clifton, G., & Nelson, J. D. (2021). A conceptual approach towards the evaluation of the vulnerability of urban railway network infrastructure by analysing railway accident reports. Australasian Transport Research Forum 2021 Proceedings. Retrieved from https://tinyurl.com/mprw4um3

Hong, W. T., Clifton, G., & Nelson, J. D. (2020). Evaluating the vulnerability of urban railway network infrastructure by analysing railway accident reports. Transport Research Association for New South Wales Symposium 2020. Retrieved from https://tinyurl.com/7vw4fn94

Statement of originality

This thesis has not been submitted for any other degree or purposes. I certify that the intellectual content of this thesis is the product of my own work and that all the assistance received in preparing this thesis and sources have been acknowledged.

Wei-Ting Hong, 1 November 2023

As supervisor for the candidature upon which this thesis is based, I can confirm that the authorship attribution statements and statement of originality are correct.

Dr Geoffrey Clifton, 1 November 2023

Professor John Nelson, 1 November 2023

Abstract

While analysis of the vulnerability of rail transport systems has received considerable attention in the literature over several decades, few previous studies offer a holistic view of how hazards and their mechanism in railway accidents make rail transport systems vulnerable in different jurisdictions and time periods. Although railway accident reports and recommendations are proposed by independent investigators to reveal relevant hazards and vulnerabilities after accidents to maintain a safer railway operational environment, practitioners and researchers suffer from the need to deal with a large amount of textual data given that most railway safety-related information is recorded and stored in the form of text. This also means there is no general model for incorporating a range of data sources and requiring only limited human intervention in the literature. Hence, there is a growing need for accurate estimations of the vulnerability of railway transport and for effective mitigation strategies.

This thesis extends knowledge on the vulnerability of the railway system by exploring the underlying hazards and building rigorous and automated models to enlarge the database by applying state-of-the-art techniques. The conceptual frameworks *HazardMap* and *RecoMap* were developed to overcome this gap, using open-sourced Natural Language Processing (NLP) topic models BERTopic and the Structural Topic Model (STM) for the automated analysis of textual data to extract critical insights. The topic modelling depicts the relationships between hazards, railway accidents and investigator recommendations and is further extended and integrated with the existing risk theory and epidemiological accident models. Empirical data was retrieved from official railway accident reports published by four countries: Australia - the Australian Transport Safety Bureau (ATSB), the UK - Rail Accident Investigation Branch (RAIB), the US - National Transportation Safety Board (NTSB) and Canada - the Transportation Safety Board of Canada (TSB). The railway accident ontology

is introduced to describe the nature of railway accidents and standardise the terminology used in different countries. Scoping workshops and a survey were conducted to evaluate the usefulness and consistency of railway practice. Case studies of the application to the risk at level crossings and the platform–train interface risks are provided to illustrate how the models proposed work with real-world data.

HazardMap shows that hazards share partly part of similar mechanisms across countries, implying that they share similar characteristics and result in similar vulnerabilities and railway accidents. On the other hand, *RecoMap* reveals that the concept of triple-loop learning is insufficient in the railway industry of the investigated countries, implying that current practices might result in railway accidents that could have been prevented by learning from other jurisdictions and implementing corresponding mitigation measures in advance. The scoping workshops and survey also revealed that the current approach primarily concentrates on jurisdiction-based analysis over time rather than learning across jurisdictions, indicating an emerging vulnerability across rail transport systems.

The interpretation of findings supplemented by additional evidence and existing theories indicates the potentially emerging hazard of deterioration in railway safety and how current stable railway systems worldwide may become hazardous. Potential barriers to learning across jurisdictions and time might deteriorate the organisational safety culture and endanger railway safety unless further strategies are implemented to stimulate the learning culture. To address such obstacles, the *HazardMap* and *RecoMap* proposed are capable of automating hazard analysis with adequate accuracy to help stakeholders better understand hazards and help practitioners learn across jurisdictions and time. Further research could incorporate data from additional jurisdictions or from earlier time periods and the frameworks could be applied to road, aviation or maritime accidents.

Keywords: railway safety, vulnerability, railway accident analysis, Natural Language Processing, ontology

Acknowledgements

As the journey of my PhD study has finally come to an end, I look back and realise how amazing, enriching and rambling this experience was. Although COVID-19 meant I was unable to attend any in-person activities at ITLS during my first and second year of study, the support of people from a variety of groups helped me to go through such a frustrating period.

First, I would like to express my sincere appreciation to my primary supervisor, Dr Geoffrey Clifton, for his valuable advice, guidance, and engagement. I would also like to express my gratitude to my auxiliary supervisor, Professor John Nelson, for his unwavering support and for providing a splendid research environment.

Second, I extend my appreciation to all members at the Institute of Transport and Logistics Studies (ITLS) at the University of Sydney for offering a warm and friendly ambience to stimulate innovation. The ITLS family is always on my side whenever I need any assistance. I would also like to appreciate the valuable contributions and time of participants in the scoping workshops and survey, providing me with in-depth insights and meaningful feedback.

I am eternally grateful to my immediate family, friends and colleagues in Taiwan, especially my parents-in-law, Pin-Hao and Liang-Su, for their dedicated support during the COVID-19 lockdown, and my mother, Chia Ling, for immense support and encouragement.

Finally, I would like to acknowledge the generous funding from the Taiwan Ministry of Education - University of Sydney Scholarship (funding code is SC3261), the Research Travel Support Scheme from the Business School at the University of Sydney, and the Postgraduate Research Support Scheme from the University of Sydney.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

Abbreviation	Full name
AHP	Analytic Hierarchy Process
ALARP	As Low As Reasonably Possible
ALCAM	Australian Level Crossing Assessment Model
ARAIB	Aviation and Railway Accident Investigation Board, South Korea
ATSB	Australian Transport Safety Bureau
BERT	Bidirectional Encoder Representations from Transformers
BNF	Backus-Naur Form
BoWs	Bag of Words
CAST	Causal analysis based on STAMP
CBOW	Continuous Bag-of-Words
CFF	Contributing Factors Framework
CG	Candidate Generation
CNN	Convolutional Neural Network
COPFS	Crown Office and Procurator Fiscal Service, Scotland
CV	Coefficient of Variance
DL	Deep Learning
DSB	Dutch Safety Board
EC	Eigenvector Centrality
ED	Entity Disambiguation
FAIR	Findability, Accessibility, Interoperability and Reusability
FRA	Federal Railroad Administration, United States
HDBSCAN	Hierarchical Density-Based Spatial Clustering of Applications with Noise
HFACS	Human Factors Analysis and Classification System
HRO	High-Reliability Organisation
ICAM	Incident Cause Analysis Method
ICAO	International Civil Aviation Organisation
IDF	Inverse Document Frequency
INPO	Institute of Nuclear Power Operations, United States
IRSC	International Railway Safety Council
IUR	International Union of Railways
JTSB	Japan Transport Safety Board
KG	Knowledge Graphs
LC	Level Crossing
LDA	Latent Dirichlet Allocation

Abbreviation	Full name
LSTM	Long Short-Term Memory
MCS	Minimum Cluster Size
MD	Mention Detection
ML	Machine Learning
MLM	Masked Language Model
MORT	Management Oversight and Risk Tree
MTO	Man–Technology–Organisation
NER	Named Entity Recognition
NLP	Natural Language Processing
NLTK	Natural Language Toolkit
NSP	Next Sentence Prediction
NTSB	National Transportation Safety Board, United States
OWL	Web Ontology Language
POS	Part-of-Speech
RAIB	Rail Accident Investigation Branch, United Kingdom
RBV	Resource-Based View
RCA	Root Cause Analysis
RDF	Resource Description Framework
RNN	Recurrent Neural Network
RSSB	Rail Safety and Standards Board, United Kingdom
SC	Semantic Coherence
SDP	Semantic Dependency Parsing
SHELL	Software–Hardware–Environment–Liveware
SRL	Semantic Role Labelling
STAMP	Systems Theoretic Accident Model and Process
STM	Structural Topic Model
STPA	System-Theoretic Process Analysis
SVM	Support Vector Machine
SWRL	Semantic Web Rule Language
TF	Term Frequency
TSB	Transportation Safety Board of Canada
TTSB	Taiwan Transportation Safety Board
UIC	International Union of Railways
UMAP	Uniform Manifold Approximation and Projection
WHO	World Health Organization

1. Introduction

This chapter introduces the thesis in five sections: the background to the research,(Section 1.1), the purpose of the thesis, derived questions and propositions (Section 1.2), the scope of the thesis (Section 1.3), limitations of the research; (Section 1.4) and the structure of the thesis (Section 1.5).

1.1 Background

Railway transportation plays an essential role in the functionality of our society, economy and environment. With the desire for sustainable transportation increasing dramatically in recent decades, the public has become increasingly interested in railway transportation for its high-quality performance and better utilisation of scarce resources (Szymula & Bešinović, 2020). However, railway accidents significantly disrupt the transportation network and cause catastrophic impacts on society, such as fatalities, injuries and economic loss. A variety of theories and frameworks are proposed in the literature to understand the mechanism of railway accidents from various perspectives: epidemiological (Peters et al., 2018), systemic (Read et al., 2021; Santos-Reyes & Beard, 2009), causation and sequencing (Wullems et al., 2013; Xia et al., 2012), and the barrier of energy (Huang et al., 2020). Many frameworks have been widely used in the railway industry and by railway accident investigation bodies. For instance, root cause analysis (RCA), the accident causation model, and systems theory are commonly used during railway accident investigations to identify the causal relations between (underlying) factors (ATSB, 2009; Dai & Wang, 2010; Kinnersley & Roelen, 2007; RAIB, 2008).

To better understand the relationship between railway accidents and their impact, there has been a rapid rise in the use of the term “vulnerability” in the context of transportation to illustrate the consequences caused by a disruptive event or a railway accident (Bates et al., 2014; Mattsson & Jenelius, 2015; Reggiani et al., 2015). Researchers have always seen vulnerability improvement as a process with three stages: pre-disruption investment, post-disruption adaptive response, and post-disruption maintenance or re-construction (Turnquist & Vugrin, 2013). Pre-disruption investment is sometimes associated with risk management, which aims to strengthen the resilience of the railway system through identifying the hazards and applying strategies to prevent the disruptions from happening (Khoudour et al., 2011). On the other hand, post-disruption adaptive response and post-disruption maintenance are to control the negative impacts and maintain operations, which mainly focus on resource allocation and optimisation. From a risk management perspective, there is a strong association between controlling risk and improving the vulnerability of a system. The nature of risk contains two core elements: probability and consequences. Risk management aims to improve safety by either reducing the probability or mitigating the consequences. Meanwhile, vulnerability management is another side of the same coin, which aims to prevent the system from being vulnerable by eliminating hazardous factors and reducing the impact. The similarity between pre-emptive vulnerability and risk assessment has also been raised in previous studies (Pant et al., 2016; Peck, 2007). The relation between hazards, risk, vulnerability, accident, consequences, and the respective domain research area is illustrated in Figure 1-1.

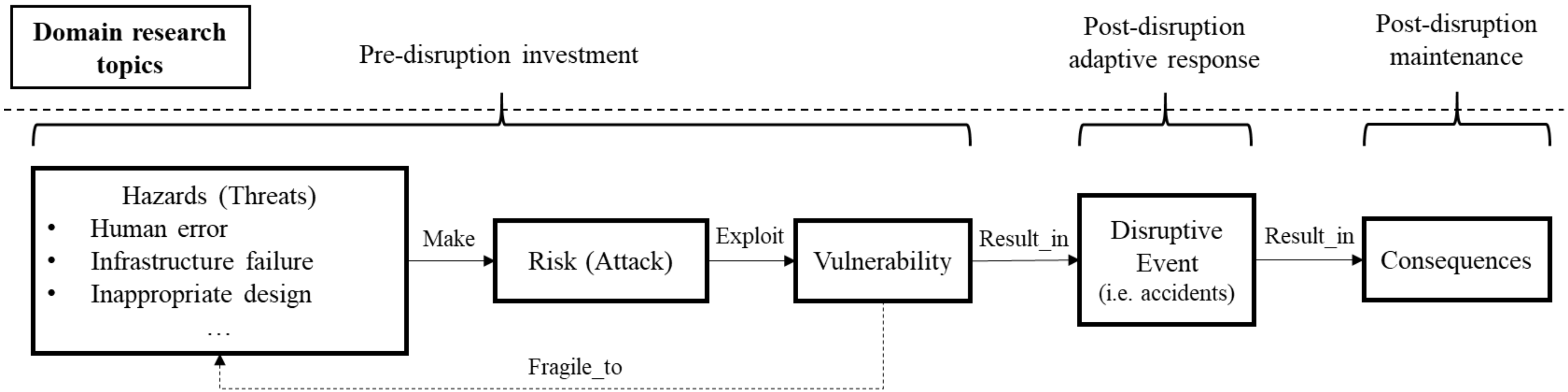


Figure 1-1: The relationship between hazards, risk, vulnerability, accident, consequences and the respective domain research topics (extended from Rausand, 2013)

Recent developments in solutions for post-disruption adaptive response and post-disruption maintenance have a number of remarkable outcomes in terms of establishing the impact (Berche et al., 2009; Chang & Nojima, 2001), optimising the resource allocation (Boorjian et al., 2012; Liao et al., 2018), constructing quantifiable metrics (Kim & Song, 2018; Wang et al., 2017) and building different scenarios (Turnquist & Vugrin, 2013; Villalba Sanchis et al., 2020). However, pre-disruption investment is still poorly understood, and most studies have explored the consequence rather than identifying the hazards and estimating the probability of which hazards trigger the accident. Some studies also argue that the preparedness of repair resources helping to reduce the disruption time should be in the scope of pre-disruption investment (Goldbeck et al., 2020). However, the foundational factors triggering an accident are not considered in such studies, and the result from studies mentioned above has failed to provide a solid conclusion on preventing the system from being damaged. Nevertheless, studies considering that issue are only in single subject areas, such as climate (Binti Sa'adin et al., 2016; Lindgren et al., 2009; Oswald & Treat, 2013) or other specific components (An et al., 2011; Ettinger et al., 2016; Lamb et al., 2019) in the rail transport system. No horizontal study has been done in this context, meaning decision-makers are disadvantaged while managing risk and mitigating vulnerability.

Our knowledge of identifying hazards and estimating the probability that each hazard triggers an accident is largely based on very limited data in the literature, as most studies have tended to focus on the impact management and resource allocation on the basis of assuming that the disruption has happened (which is also known as condition vulnerability (Dehghani et al., 2014; Kim & Yeo, 2017). However, the knowledge to point out where hazards are in the railway system and how to prevent an accident from happening is still unsatisfactory from the perspective of operation and management. The railway safety and vulnerability require more improvement to provide stable and high-quality transport services.

1.2 Purpose of thesis

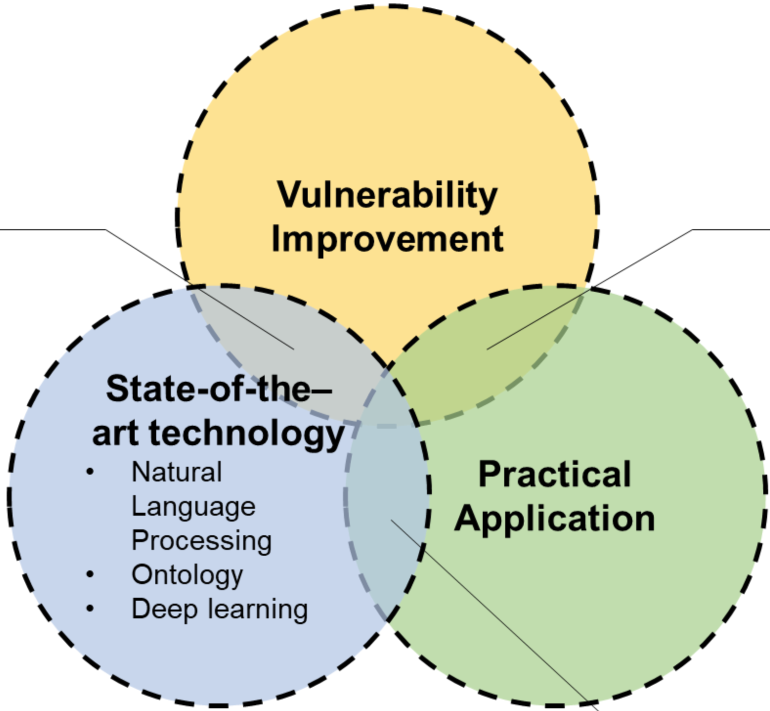
The motivation of this thesis is to evaluate and improve the vulnerability of railway networks by introducing state-of-the-art techniques to analyse the legacy accident reports and extract critical insights to meet the needs of the railway industry. Although the current railway industry has implemented several strategies to reduce the vulnerability and reinforce prevention by importing novel technology (e.g., remote condition monitoring systems and inspection systems for railway wheelsets) and the process of risk management (e.g., risk management planning, risk response planning and risk monitoring and control (Patil et al., 2008)), railway accidents still occur with unexpected hazardous factors arising from either new applications or undetected hazards. Additionally, the effects of improvement strategies are seldom reviewed horizontally. The experience gained from historical accidents is rarely shared across jurisdictions in industry and academia (Nash, 2008), meaning significant investment does not have a widespread benefit. Furthermore, railway safety improvement strategies are often implemented after railway accidents occur, but coherent pre-disruption investment should be considered beforehand, and the hazards should be identified and controlled for a strong railway safety culture.

To summarise, current solutions to improve railway safety are unsatisfactory and as they only concentrate on a particular case and make improvements based on that single case. It is noteworthy that the content of previous accidents is mostly recorded through text instead of numeric data, making them unsuitable for statistical analysis and making it difficult to extend horizontal knowledge in this context. Even in the context of academia, techniques to solve such issues are manually demanding and time-consuming due to the availability and analysability of the data. In practice, the poor communication between railway accident investigation bodies is seldom discussed and examined, and hazards might have been prevented if experience could be shared from one jurisdiction to another. For instance, the railway driver–machine interaction was widely discussed in Europe in the early 2000s, and several improvements were introduced to avoid providing overloaded information and fault alarms (Panou et al., 2013; Young et al., 2006). On 23 October 2018, a passenger train derailment occurred in Taiwan in Asia. One of the causes of this accident was driver ignorance of alarms, as multiple alarms were raised at the same time resulting in driver unawareness of the speed limit. Similar accidents have occurred in Europe before, such as the derailment on 20 February 2010 in Leicestershire in the United Kingdom. This suggests that many accidents with high similarities occurred in various countries and could have been prevented by learning across jurisdictions and over time.

This framework establishes the relation between the hazards identified from the original accident data and recommendations made to address these hazards. Some studies have raised the issue of interface failure context, most of which mainly consider a case study (Darroch et al., 2016) or only a conceptual framework (Kelly &Berger, 2006; Shokri et al., 2012). The lack of input data might result in insufficient evidence and weaken the conclusion, which can be considered other hazards in terms of railway safety improvement.

The identified research and practical gaps aforementioned are illustrated in Figure 1-2. This thesis seeks to address how to solve the present issues from the perspective of vulnerability improvement, state-of-the-art techniques, and practical applications. By combining all elements in Figure 1-2, the knowledge on railway vulnerability and risk management is expected to be broadened significantly.

Although it seems the previous study has successfully analysed enormous accident reports manually, which is large enough to do further statistical analysis. However, such a labour-intensive analysis method would have its limitation once the database is too big to read manually. Nevertheless, the framework of the model would be limited due to the present analysis method, resulting in constraints on duplicability and updatability.



Regardless of plenty of studies have discovered several elements which would trigger a railway accident individually, such as human error, technical or design failure, and environmental impact. However, none of them have looked at railway accidents from the perspective of interface failure. Furthermore, the degree of execution of recommendations has not yet been assessed, and horizontal analysis is required to reveal the real risk considering the exposure and probability.

The recommendation each railway accident investigation body made would face difficulty to consider the accidents in other jurisdictions. The gap between investigators and the railway industry or investigators per se might create unexpected risk.

Figure 1-2: Identified gaps and the respective explanation

To overcome gaps elaborated in Figure 1-2, some studies have leveraged the benefits of Natural Language Processing (NLP) and machine learning to consistently analyse a large body of textual data. NLP is the state-of-the-art technology that addresses the interface between human languages and computers by enabling a computer program to process a large amount of textual data through machine learning approaches. Several attempts to incorporate NLP into accident data analysis can be found in maritime, aviation and road safety to analyse crowdsourced textual data (Nelson et al., 2020; Syeda et al., 2019; Wang et al., 2017). Despite the extensive discussion of automated textual data analysis in the literature, the focus is mainly on building the NLP model rather than interpreting the result. Additionally, most studies in this context use the supervised learning approach, requiring a significant amount of manual effort to train the model (Sizov & Öztürk, 2013; Wang et al., 2017). These limitations hinder researchers and practitioners from advancing the existing railway safety knowledge with the help of novel technologies.

Therefore, the research objective of this thesis is to provide a holistic view of the nature of hazards in railway accidents and responses from railway industries across jurisdictions and across time by leveraging the power of NLP with little manual effort. Instead of establishing a customised model, this study only uses open-sourced and off-the-shelf toolkits for building the NLP model so that the result and contribution of this study can be duplicated and reused. The proposed models are implemented to analyse railway accident reports and recommendations, which are expected to offer another view on railway accidents from the hazard-centred perspective and advancing railway safety knowledge by enabling learning across jurisdictions and across time.

To ensure the identified research gaps are well addressed in further analysis, several research questions and propositions are derived on the basis of the research purpose and shown in Table 1-1. Three research gaps are identified with corresponding research questions and propositions, reflecting issues that might be overlooked in the literature. To sum up, this study argues that some hazards might share similar characteristics and attributes regardless of countries and jurisdictions (RQ1-1, RQ1-2) and result in similar vulnerabilities in different jurisdictions and times (RQ1-3). This implies that learning across jurisdictions might be able to mitigate hazards which have triggered accidents in other countries. Additionally, it is assumed that recommendations made by railway accident investigators might primarily focus on addressing operational issues and concentrate less on risks at the management level (RQ2-1), indicating the possibility of overlooking the importance of learning across jurisdictions and time (RQ3-1). Although there might be a transition in the style of making railway accident recommendations in each jurisdiction over time as the awareness of safety culture increases (RQ2-2), barriers to the railway industry learning across jurisdictions and time still exist (RQ3-2) and result in unexpected hazards (RQ3-3).

Table 1-1: Derived questions and propositions based on research gaps

Research gap	Research questions	Proposition
Limited understanding and comparison of hazards and their mechanism in railway accidents across jurisdictions	**RQ1-1:** What is the difference in roles each hazard plays in various jurisdictions during railway accidents?	Regardless of hazard taxonomy, certain hazards have similar attributes and have resulted in comparable railway accidents across different jurisdictions.
	RQ1-2: Do the same hazards occur in different jurisdictions and across time?	There are some hazards sharing similar characteristics and occurring in different jurisdictions and across time.
	RQ1-3: Do those hazards result in similar vulnerabilities in different judications and times?	Those hazards sharing similar characteristics and occurring in different jurisdictions and across time may result in similar vulnerabilities and railway accidents.
Limited understanding of how railway accident investigators in different jurisdictions address hazards identified over time	**RQ2-1:** How do recommendations made by railway accident investigators address hazards identified from the socio-technical perspective?	Recommendations made by railway accident investigators might primarily focus on addressing operational issues and concentrate less on risks at the management level.

Table 1-1: Derived questions and propositions based on research gaps (continued)

Research gap	Research questions	Proposition
	RQ2-2: Is there a transition in the style of making railway accident recommendations in each jurisdiction over time?	The style of making railway accident recommendations might change over time, resulting in a potential change in the way that recommendations are proposed.
Limited understanding of the relationship between the learning behaviour in the railway industry and its impact on railway safety culture	**RQ3-1:** Do railway accident report recommendations support the railway industry to learn across jurisdictions and time?	Recommendations made in railway accident reports often overlook the importance of learning across jurisdictions and time.
	RQ3-2: What are potential barriers to the railway industry learning across jurisdictions and time?	The railway industry has multiple barriers to learning across jurisdictions and time in the socio-technical hierarchy.
	RQ3-3: What hazard(s) might emerge if barriers to learning across jurisdictions and time remain unsolved?	The absence of learning across jurisdictions and time might significantly impact safety culture.

1.3 Scope of thesis

On top of that, this thesis aims to shed new light on the vulnerability of railway systems by exploring the underlying hazards and building a rigorous model to enlarge the database by applying state-of-the-art techniques. Current work mainly concentrates on reducing the vulnerability of the railway system by preventing railway accidents from happening, which is also known as the pre-disruption investment in the literature (Hua & Ong, 2017; Pant et al., 2016). By following this approach, the complexity in major railway accidents should be decomposed in a logical way, and horizontal analysis conducted for the purpose of reducing the prevalence of accidents and contributing to enhanced learning between railway accident investigation bodies in different jurisdictions.

Before investigating the critical components in pre-disruption investment, it is important to clarify what can be referred to as a railway accident. The term “railway accident” is defined as an accident or incident which occurs on railway property in so far as it is or may be relevant to the operation of the railway (RAIB, 2005); however, not every railway accident results in disruption and increases the vulnerability of the whole system. Therefore, the thesis focuses on railway accidents resulting in one of the following five outcomes: fatalities, major injuries, derailments, collisions, and other specified dangerous occurrences, such as the emission of toxic liquid. Those railway accidents containing one of these five outcomes are among the most commonly discussed and well-documented. Most importantly, they create an extreme impact on society, the economy, and the whole railway network, and the hazards triggering such railway accidents make the whole system vulnerable to operate.

The scope of this thesis is the interactions between hazards, railway accidents, rail accident investigation bodies, railway industries and information and knowledge related to railway safety. The definitions of terms mentioned above are discussed in Section 2.1. The information and knowledge flow, railway organisations of interest and their relations in the railway system and the research questions investigated in this study are shown in Figure 1-3. There are different hazards across countries triggering railway accidents, resulting in potential knowledge and information for railway industries to understand and analyse. Rail accident investigation bodies lead the investigation and generate knowledge (in the form of recommendations) for improving railway safety through organisational learning within country A. Despite no direct impact, similar railway accidents that have occur in other countries and knowledge generated might be valuable information and knowledge for the railway industry in country A. Such learning behaviours can be conducted by rail accident investigation bodies or railway organisations. Although there are several relations worth investigating in Figure 1-3, this study only focuses on the relationship between hazards, railway accidents, knowledge generated by rail accident investigation bodies and the learning behaviour of railway industries.

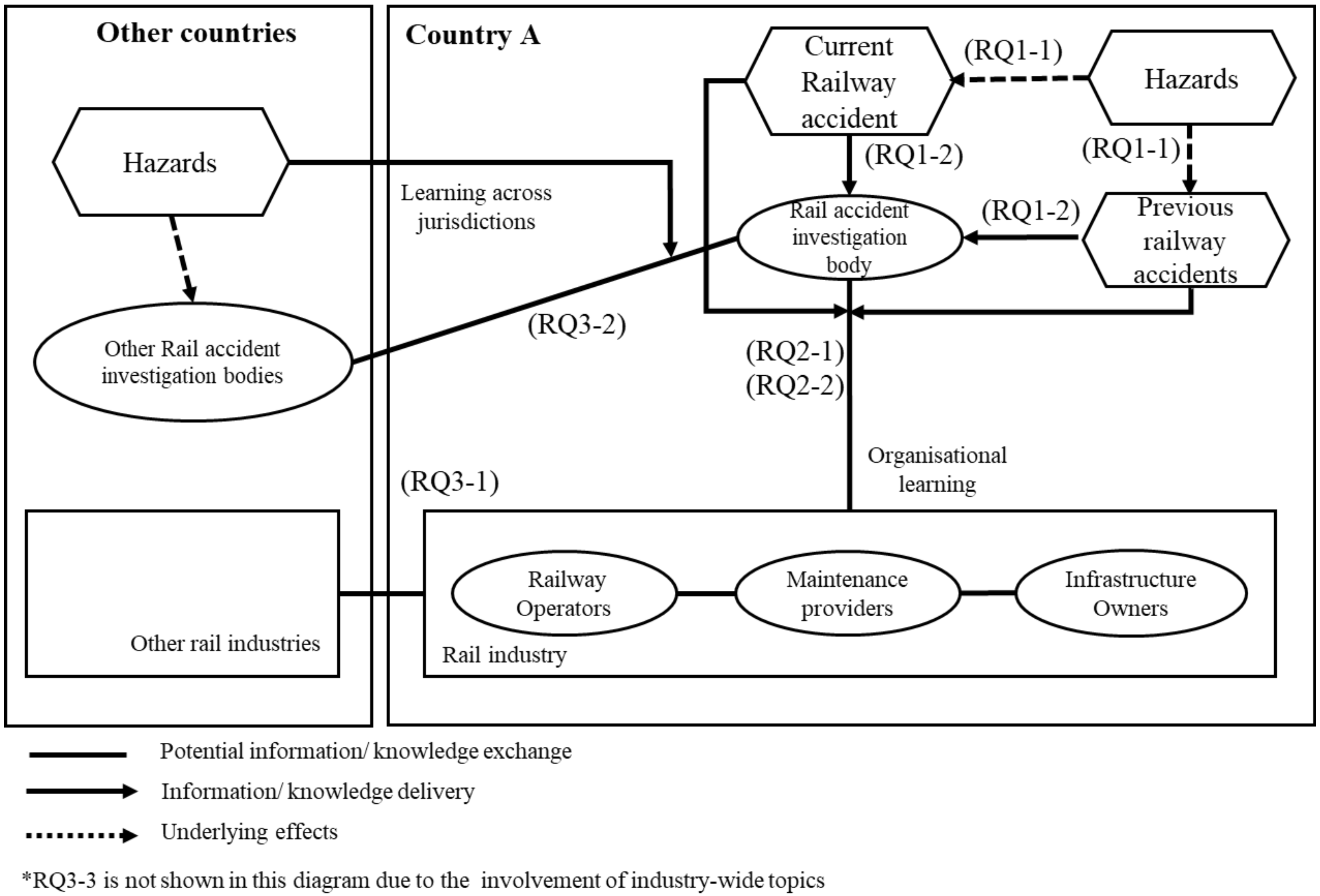


Figure 1-3: The information and knowledge flow, railway organisations of interest and their relations in the railway system and the research questions investigated in this study

Multiple data resources are considered in this thesis to examine whether experience is shared effectively in the railway industry. However, the difference in railway systems, language in recording accidents, culture, and regulations would make the analysis inconsistent and create significant bias. Hence, only countries where investigation bodies have the following features are considered:

(1.) The investigator must have a comprehensive documentation system to reduce the complexity of processing. The framework of the accident report must be clear and consistent in the temporal aspect. For instance, the jurisdiction has a law or regulation on the format of generating accident reports.

(2.) The investigator must have been granted the independent authority to conduct the investigation. The investigation objective should aim to increase railway safety regardless of blame or liability.

(3.) The investigator should have a foundational classification system regardless of how simple it is, to make it easier to merge data from different resources.

(4.) The investigator must produce reports that contain recommendations, which focus on issues relating to railway safety, such as the implementation of specific training or policy, introducing new technology, or revising existing standard operating procedures. The recommendations must not contain inferences or conclusions of apportioning liability.

(5.) Due to the limitation of the technique applied in this research, only data from native English-speaking countries is considered to ensure the performance of the model, which implies that reports must be written in English. The English language used in reports should be consistent regardless of time, the types of accident, or investigation engagement. For instance, the definition of derailment in each jurisdiction should not vary through time.

(6.) Due to the differences in the use of the English language, the model requires sufficient data to modify the parameters on the basis of the terminology each jurisdiction uses. Therefore, to ensure the performance of the model this study only considers investigators that have completed over 100 reports.

Based on these requirements, four countries are considered, namely Australia, Canada the United Kingdom (UK) and the United States of America (US).

1.4 Structure of the thesis

This thesis is organised as follows:

Following an introduction to the background, purpose and scope of this research, Chapter 2 presents a comprehensive literature review in the context of vulnerability assessment, risk management, and the analysis of railway accident reports. The terminology is defined, and the boundaries of current studies are explored. An overview of how literature handles the issues of railway vulnerability and the gained results are presented and analysed. The analysis of railway accident reports in the literature and practice is reviewed. Last, the synthesis of findings is presented and the research gaps are confirmed.

Chapter 3 reviews different approaches used in the analysis of the aforementioned contexts and discusses the research design of this thesis. The mechanism of NLP and its novel techniques are introduced, including approaches and resources required for training NLP models. Last, the concept of ontology design and knowledge graphs for practical applications in the literature are discussed.

Chapter 4 elaborates on the development of the models. First, the framework of NLP models and applied data is introduced. Candidate approaches are compared. Second, entity linking strategies are designed to standardise terminology used in different countries. The development of ontology knowledge graphs used to address the entity linking task is depicted.

Next, the covariate analysis method provided by NLP models is discussed and demonstrated. Last, details of the scoping workshops and survey used to evaluate the models developed are presented.

Chapter 5 and Chapter 6 concentrate on the analysis and the discussion of findings, starting with data acquisition and data pre-processing. Next, the analysis results of each country are described in detail, including the initial outcomes, model fine tuning, performance evaluation and justification. Subsequently, details of ontology, knowledge graph selection and entity linking are revealed with the example. Cross-sectional analysis of countries and investigators is conducted followed by the scoping workshops plan for evaluation. Last, *HazardMap* and *RecoMap* are proposed with the case study of level crossing accidents and the platform–train interface risk.

Chapter 7 synthesises findings and discussion of previous chapters and further extends the discussion of potential implications based on the result. Learning behaviours in the railway industry are discussed, followed by knowledge flow within and across railway industries. A concept of an emerging underlying hazard is proposed indicating the potential deterioration of railway safety culture. The opportunity to overcome barriers mentioned above is discussed.

Chapter 8 summarises findings, highlights contributions and limitations of this thesis and makes suggestions for future research.

2. Literature review

This chapter presents a comprehensive review of current literature. First, there is a brief discussion of the historical context of vulnerability in the field of transportation research (Section 2.1). Second, popular assessment approaches in the field of analysing vulnerability are demonstrated (Section 2.1). Approaches relating to the context of the railway are also highlighted in this section. The next section concentrates on how a railway accident report is organised and the linkage to the evaluation of the vulnerability of a railway system. Jurisdictions are compared in how they investigate a railway accident and how they produce a report is included (Section 2.3). The frameworks applied to assess railway-related risk in the literature are introduced (Section 0). Last, the literature is synthesised, and the scope of this study is determined.

2.1 The concept and elements of vulnerability

To build a vulnerability analysis system for a railway transport system, several closely related topics are defined: *vulnerability*, *resilience*, *robustness* and *reliability.* These are popular keywords used in the context of vulnerability assessment literature. Typically, the definition of these keywords is modified based on the features or characteristics of the research. To discriminate between them, *vulnerability* is defined first followed by comparison with other keywords.

Vulnerability is commonly defined as “the sensitivity of a system to threats and hazards” regardless of the research area (Rausand, 2013), such as evaluating the vulnerability of electric power delivery networks (Holmgren, 2004), building a risk measurement framework for vulnerable infrastructure (Haimes, 2006), and assessing the vulnerability of a road traffic system to guide maintenance priorities (Jenelius et al., 2006). Additionally, in order to do further numerical research, a group of studies has created a common awareness of a definition of vulnerability from a quantitative perspective as in the following description: how much performance is diminished during a disruption (Berdica, 2002; Bešinović, 2020; Khaled et al., 2015; Yu et al., 2018; Zhang et al., 2018; Zhou et al., 2019). The meaning of performance varies depending on the research topic. For example, it can refer to the cost of travel when considering a road transport network (Jenelius et al., 2006), headway when in a railway transport network (Cadarso et al., 2013), and delivery time in a freight logistic network (Khaled et al., 2015).

Sometimes the term *vulnerability* is slightly modified in response to a specific scenario. For instance, Laurentius (1994) and Berdica (2002) deem that vulnerability is a susceptibility for rare, big risks or the sudden, unpredicted occurrence, while Rausand (2013) suggested that vulnerabilities appear when a system begins to be destroyed until it totally collapses following small further stresses when the capacity reaches its maximum: “Little strokes fell great oaks”. The disadvantage of considering the probability of failure while analysing vulnerability is revealed by Sarewitz et al. (2003), who argue that the estimation of probabilities in the analysis of extreme events is impractical, and then propose the idea of *conditional vulnerability* (or *exposure* in other literature, see D’Este & Taylor (2003) and Jenelius et al. (2006)), which is used to calculate the aggregate result of consequences given a hazardous event occurs.

Second, *resilience* has been commonly defined as “the capability to persist and absorb a disruption without affecting performance” in the context of general transport systems (Mattsson & Jenelius, 2015) or a specific transport system, such as a railway (Bešinović, 2020). A slightly different operational definition of *resilience* has been made regarding the resources that a system requires to re-strike its balance between demand and supply (Zhou et al., 2019). When the research is considering the resilience of the socio-technical system like a healthcare system, then the objective would be the ability to absorb a disruptive event or the time it takes to recover from shocks of disruptive events (Taysom & Crilly, 2017). On the other hand, when the subject switches to a transport network, then the performance to maintain functionality under disruptions or time and resources required to recover to normal status would be the focus. Additionally, if the research considers the temporal factor in resilience analysis, it might be referred to as *resilience engineering* research. Resilience engineering research is a paradigm for safety management, which offers a much broader socio-technical framework to cope with infrastructure threats and disruptions, including preparedness, response, recovery and adaptation (Patriarca et al., 2018; Worton, 2012).

Another term that is similar to *resilience* is *robustness*. There is an inconsistent definition of robustness. For instance, Bešinović (2020) considers that in the context of the railway industry, robustness is the ability to mitigate impacts caused by disturbances, whereas resilience means the ability of a system to provide adequate service in normal conditions. However, Zhou et al. (2019) argue that robustness is the ability of a system to maintain itself in its original state regardless of the impact which the disruption creates; and once the impact makes a system unable to maintain normal status, then resilience would be considered as the main topic. Both of these studies support the idea that if an operational index is needed to evaluate the remaining level of performance under disruptions, then resilience research should be considered. But when it comes to robustness, Bešinović (2020) assumes the performance has been affected and regular service might not be provided, whereas Zhou et al. (2019) assume the performance remains the same during disruption.

The apparent contradiction in definition might be caused by the different extent of disruption. Zhou et al. (2019) argue that the use of both the terms robustness and resilience depends on whether the network can provide the same performance while disrupted. If an event occurs and a transport network can still provide the same level of service, then the event is a *disturbance* (Nielsen, 2011) and *robustness* can be used in this context. On the other hand, if an event happens and degrades a transport network , the event is referred to as a *disruption* (Nielsen, 2011) and the term *resilience* is used to estimate the impact of the event as a result. Note that resilience is considered from the aspect of performance or capability rather than the time a network takes to recover to normal status. In the context of time aspect, *dynamic resilience* (or namely *rapidity*) is used to describe the time that a system requires to return to a state of normal function after a severe perturbation, such as after an intentional terrorist attack (Pimm, 1984; Wang et al., 2016).

The terms mentioned above can be shown diagrammatically, as proposed by McDaniels et al. (2008), in Figure 2-1.

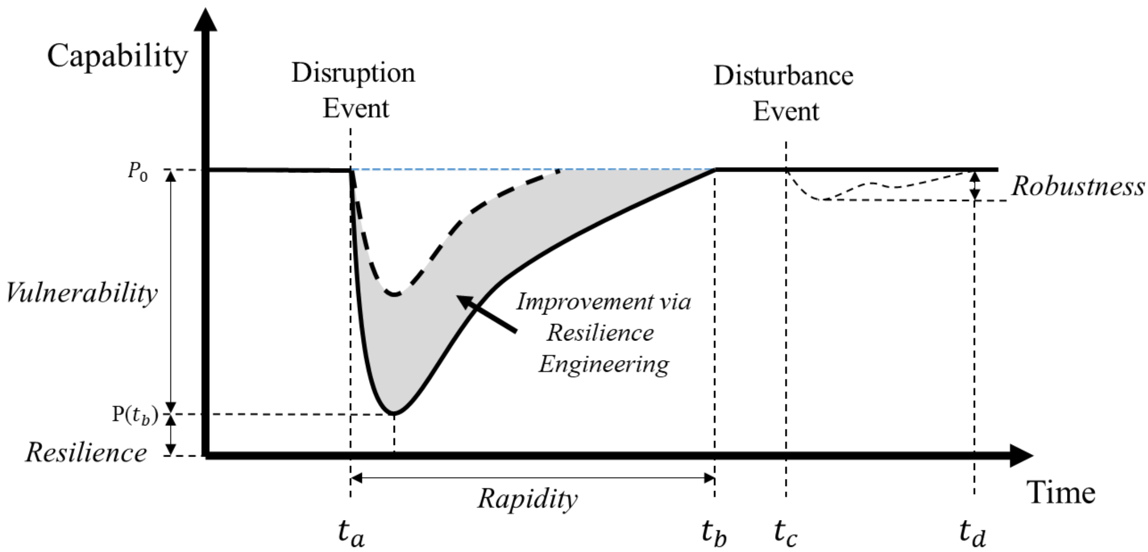


Figure 2-1: Relationship between vulnerability, resilience and robustness, revised from McDaniels et al. (2008)

Assume P_0 is a regular capacity of a transport system network, a disruption (e.g., terrorist attack) occurs at time t_a and reduces the capacity to the most disrupted point $P(t_b)$ at time t_b. The capacity gap between P_0 and $P(t_b)$ is referred to as the vulnerability of the network. In addition, the rest capacity (P_0 - $P(t_b)$) is referred to as the resilience of the network. The vulnerability in a network stands for the susceptibility to a disruption resulting in a considerable reduction, whereas the resilience in a network stands for the rest capacity after the degradation due to a disruption (Berdica, 2002; Mattsson & Jenelius, 2015; Szymula & Bešinović, 2020). On the other hand, when a disturbance occurs at time t_c, the network absorbs the disturbance, and the capacity remains the same as usual. The potential loss of capacity absorbed by the network is robustness. The grey area derived from the dotted line and solid line implies that if resilience engineering is conducted to help improve the network when disruption is persisting, then the degree of impact on the aspect of rapidity and vulnerability will decrease. For instance, if the design of the rolling stock could absorb more impact which is caused by an explosion and prevent a derailment, then the cost of time will be reduced dramatically because no heavy machinery is required to handle the disruption.

Reliability can be interpreted from the perspectives of demand and supply (Jenelius et al., 2006). From the demand side, Taylor and D'Este (2007) made a clear distinction between reliability and vulnerability in a network. They considered that reliability focuses on connectivity, which refers to the probability that a specific trip between an origin–destination pair can be completed within a particular time. At the same time, vulnerability reflects the consequences (often measured by an index of accessibility) of the loss of a particular number of nodes or links in a network. Immers et al. (2004) defined reliability as the degree of certainty with which a passenger is able to estimate their travel time. From the demand side, reliability can be assessed by estimating the performance of a transport network. However, measuring the probability of disruptions has long been a challenge for researchers. For instance, the probability that a passenger finishes their trip in an estimated time, also known as *travel time reliability,* is difficult to estimate accurately (Bell & Iida, 1997; Taylor & D'Este, 2007). Another example is the probability that a network can successfully accommodate a given level of travel demand, also known as *capacity reliability* (Yang et al., 2000). Some studies claim that the issue of probabilities should be concerned more than consequences (Berdica, 2002;

Jenelius et al., 2006), while others choose to deal with this issue on the supply side regardless of probabilities.

Mattsson and Jenelius (2015) provided a review of recent research and purely defined reliability in a theoretical way: the probability that there is still a path between a pair of nodes when one or more links or nodes are removed. Zhou et al. (2019) considered that reliability stands for the ability of a network to maintain its performance under the condition of disruption, which is similar to the previous statement about the term resilience. However, in either approach, the definitions all assume that a disruptive event happens and calculate the consequence as the proxy of reliability in a transport network.

The main highlights and definitions of disruption in railway systems inferred through literature are summarised in Table 2-1.

Table 2-1: Definition of terms relating to disruption in a railway system

Terms	**Definition**	**Source**
Vulnerability	Descriptive definition: Sensitivity to threats and hazards, or a susceptibility to incidents that can result in considerable reductions of performance in a system.	(Berdica & Mattsson, 2007; Taylor & D'Este, 2007)
	Operational definition: Based on the descriptive definition, it refers to how much performance is taken by a disruption.	(Berdica, 2002; Bešinović, 2020; Khaled et al., 2015; Szymula & Bešinović, 2020; Zhang et al., 2018; Zhou et al., 2019)
Conditional vulnerability (Exposure)	The aggregate consequences of a given hazardous event to a system.	(D'Este & Taylor, 2003; Jenelius et al., 2006)
Disruption	An event or a series of events caused by external or internal factors that leads to substantial deviations from planned operations. From the perspective of risk management, a disturbance event turns into a disruption after it begins to lead to deviations.	(Nielsen, 2011; Zhou et al., 2019)
Disturbance	A disturbance indicates an event that causes part of the railway operations to deviate from the operational plans, but the system can still provide the same level of service as usual.	(Nielsen, 2011; Zhou et al., 2019)

Table 2-1: Definition of terms relating to the disruption in a railway system (continued)

Terms	**Definition**	**Source**
Reliability	<u>Demand side:</u> Travel time reliability: The probability that a specific trip between an origin–destination pair can be completed within a specific time. Capacity reliability: The probability that a network can successfully accommodate a given level of travel demand.	(Immers et al., 2004; Taylor & D'Este, 2007)
	<u>Supply side:</u> The probability that there is still a path between a pair of nodes when one or more links or nodes are removed.	(Mattsson & Jenelius, 2015)
Resilience	<u>Descriptive definition:</u> The ability of a transport system to prepare for and to withstand, absorb and adapt to shocks, and to recover from the consequences in a timely and efficient manner under disruptions.	(Berdica, 2002; Mattsson & Jenelius, 2015)
	<u>Operational definition:</u> The remaining capacity after the degradation due to a disruption.	
Dynamic resilience	The rapidity with which a system returns to a state of normal function after a severe perturbation.	(Pimm, 1984; Wang et al., 2016)
Resilience engineering	A subject offering a much broader socio-technical framework to cope with infrastructure threats and disruptions, including preparedness, response, recovery and adaptation.	(Worton, 2012)
Robustness	The capability to withstand disturbances with an acceptable reduction in operating performance, measured by the potential decrease of capacity.	(Pagani et al., 2019; Zhou et al., 2019)
Rapidity	The speed of recovery to normal status.	(McDaniels et al., 2008)
(un)satisfied demand	If passengers need to wait for a train more than 5 minutes more than they do in a normal situation, then it is defined as an unsatisfied demand.	(Piner & Condry, 2017)

2.2 Vulnerability assessment

Since transport operators generally agree that recurring disruption is inevitable and usually followed by a series of consequences, assessing the vulnerability of a transport network has attracted increasing focus. This section reviews the theories proposed in the field of vulnerability analysis. The approaches applied in the context of vulnerability assessment are introduced. Generally, the methods used in literature can be broadly classified as the statistical approach, topological analysis, simulation and optimisation, and risk analysis (Bešinović, 2020; Zhou et al., 2019).

2.2.1 Foundational theories applied in vulnerability assessment

Before 1970 the concepts of vulnerability assessment and risk assessment were mainly treated in a qualitative way (Rausand, 2013). The quantitative approach began to be widely applied in this context after 1970 due to the development of *probability theory*. Probability theory allows researchers to express the risk of an accident in a specific situation as a number between 0 and 1 (Parzen, 1960). Another advantage of applying probability theory to risk assessment or vulnerability assessment is that the frequency and the condition of an event can be clearly displayed by using the same symbols and formulas. For example, the nature of uncertainty of an accident was well demonstrated through probability theory (see Apostolakis, 1990; Karimi & Hüllermeier, 2007, Silva et al., 2008) and subsequent applications used Bayesian network models and fault tree analysis.

Since risk research has its origins in the qualitative perspective, one of the most widely used theories in this context was *reliability theory* (Bazovsky, 1961). Reliability theory aims to use either statistical approaches or mathematical models to manage the risk. Additionally, in order to optimise the performance of the systems, research in this context considers not only the purpose of the system, but also the economic context during operation (Barlow & Proschan, 1996; Pieruschka, 1963). However, reliability theory is a practice-oriented theory, which concentrates on the probability and the result of an accident. Compared with the identification of hazards, reliability theorists care more about how to estimate the reduction of ability to operate and repair (Barlow & Proschan, 1975). The interactions between hazardous elements, which might be the causes of an accident in a complex system, are less considered.

To gain better understanding of the complexity of the interaction of hazards, the concept of system accident is proposed, which is also known as the *normal accident theory* (Perrow, 1984). Normal accident theory assumes that a complex system (or socio-technical system) encounters accidents naturally and inevitably, and Perrow (1984) concludes two characteristics, interactive complexity and tight coupling, make a complex system sensitive to accidents. Interactive complexity means the interaction between more than two unexpected failures, and tight coupling means that each component in a system cannot be isolated or shut down independently due to the connection among them (Perrow, 1984; Rausand, 2013). Given the accident is inevitable, Perrow (1984) argues that the analysis should "focus on the properties of systems themselves, rather than on the errors that owners, designers, and operators make in running them". In doing so, normal accident theory reckons that the interaction between elements in a complex system should be emphasised to increase the reliability.

Normal accident theory has been heavily criticised that it does not provide any metrics to evaluate the interactive complexity and tight coupling (Sammarco, 2003), and for the underestimation of human error (Shrivastava et al., 2009; Vaughan, 1999). However, it still provides a distinct framework on organisational structure and situations during the period of an accident (Shrivastava et al., 2009). A distinct chart can be generated based on the interaction and coupling derived from normal accident theory, which helps classify industries depending on the complexity. A chart designed by Shrivastava et al. (2009) is shown in Figure

2-2. Rail transport is positioned in Quadrant 1, which represents a system which has linear interactions and tight coupling.

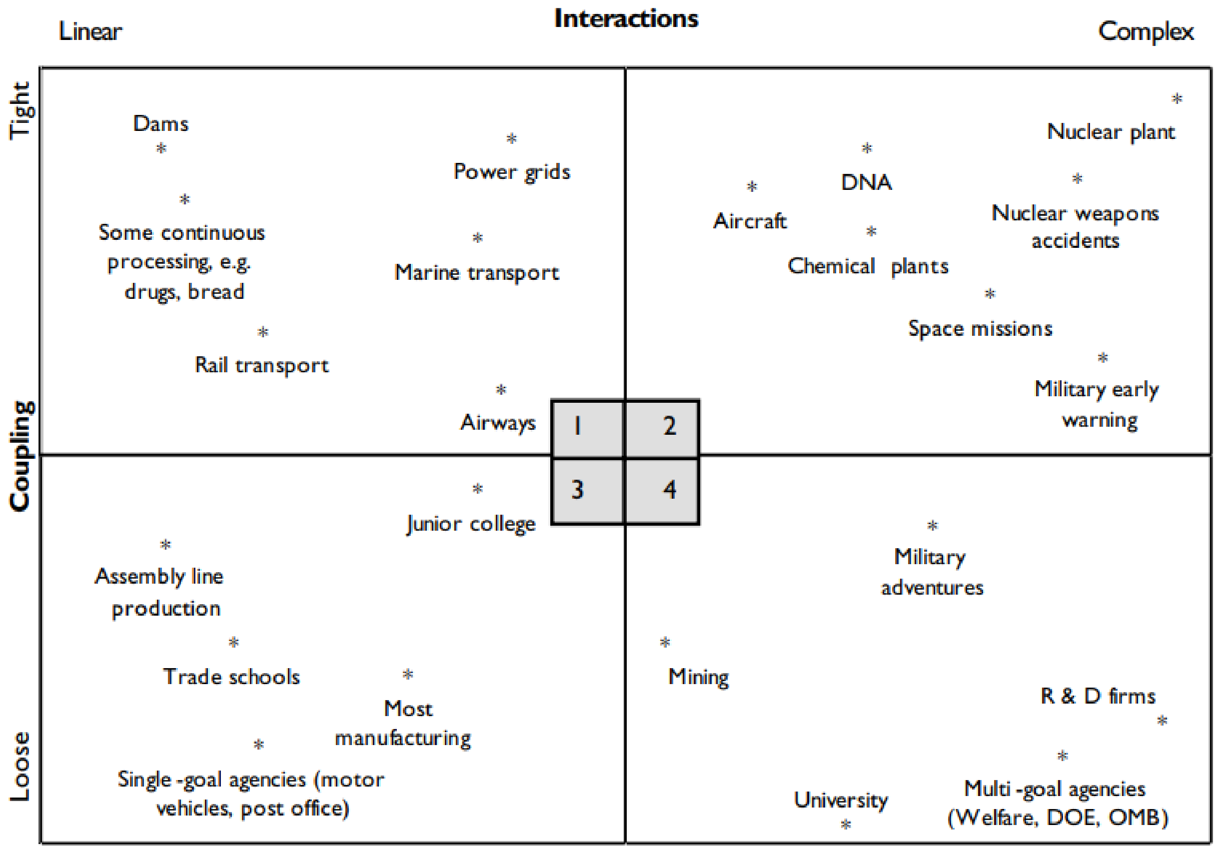


Figure 2-2: Interaction/coupling chart, source: Shrivastava et al. (2009)

Another popular theory which explains the complexity of systems is *systems theory*. Systems theory claims that every system has its own structure and purpose, and is influenced by the environment, other systems and even itself. The concept of systems theory assumes each technology would develop independently at first, but that all technologies would overlap with civil systems, social systems or technologies *per se*. A hierarchy of levels of organisations or systems would build up the complexity of the socio-technical system (Miles Jr, 1973). In terms of the application in the context of risk assessment or vulnerability assessment, the idea of the imposition of constraints and control loops in a system is widely considered to build the model. For instance, Leveson (2004) used systems theory as a foundational basis to build a model for hazard analysis and accident prevention strategies. Moreover, a popular risk analysis model called systems theoretic accident model and process (STAMP) is derived from systems theory (see Section 6.1). In contrast to normal accident theory and reliability theory, systems theory can interpret the epidemiology of accidents in modern society comprehensively.

2.2.2 Vulnerability assessment – statistical approaches

The aim of the statistical approach is to derive a feasible model to interpret the data and highlight insights into historical disruptions. Three methods are mainly used in this context: *data-driven approach*, *probabilistic risk assessment* and *experts' judgment system*. Each is briefly introduced.

First, the *data-driven approach* uses historical data directly and analyses the data before use as a performance indicator (Bešinović, 2020). However, such a method requires sufficient data, there are only a few studies, and they are rare in a railway industry context. Hong et al. (2015) applied the multi-step methodology (Monte Carlo simulation) to estimate vulnerability on the basis of 30-year flood data in China, which is popularly considered a novel approach to estimate the vulnerability of a transport network. Diab and Shalaby (2020) used data from 12,500 detailed metro rail incidents in 2013 from the City of Toronto, Canada. They built a linear and logistic regression model and identified that open-air tracks are four times more vulnerable than covered tracks in terms of number of delay incidents and total delay time. However, the availability of disruption data is extraordinarily difficult to collect (Alexakis et al., 2014). Additionally, the statistical approach usually misestimates the probability of a rare disruption.

Second, *probabilistic risk assessment* assumes the probability of a disruption follows a specific distribution. For instance, Cats et al. (2016) assessed public transport network risk analysis considering exposure of the disruption. By assuming that frequency follows Poisson distributions and applying topological analysis, it is concluded that the vulnerability of a transport network should be estimated by the reduction of passengers' welfare caused by the frequency and duration of the disruption. This approach is usually critiqued due to the inability to validate the fitness of distribution.

Finally, the *experts' judgment system* assumes that the actual probability of a disruption can be measured through objective data rather than subjective data based on the following reasons (An et al., 2011). First, the statistical methods are not able to handle the uncertainty of information. The insufficient number of disruptive events usually leads to significant bias when estimating the probability. Second, the statistical data does not exist or is not recorded. Even if the data is available, it is hard to analyse due to lack of crucial information, fragmentation, inconsistency or a high level of uncertainty. Last, the existence of data relating to disruptive events can contradict the reputation of operators. Almost all disruption data is generated and recorded by operators themselves instead of third-party organisations. This makes disruption data not only confidential but also suspect. Many studies successfully estimate the probability of a disruption by conducting expert assessment. For example, Alexakis et al. (2014) overcame the difficulty relating to the issue of probability by applying expert assessment and the cellular automata Markov model to evaluate and predict the probability of landslides in Cyprus and An et al. (2011) determined the relative importance of the risk contributions in the context of the railway industry through expert judgment.

Nevertheless, the issue of how to evaluate the probability of a disruption is gradually emerging; however, there is no known study which tries to analyse disruption in a microscopic scale. Even though some studies consider the probability when analysing the risk of a disruption instead of conditional vulnerability in the context of the railway industry, there is no research on analysing the causality between infrastructure and disruptions deeply. Hence, this thesis applies state-of-the-art techniques to overcome such difficulties. By filling this research gap, through the research in this thesis, we will be able to better understand how a disruption occurs and try to prevent and even predict disruption in the future.

2.2.3 Vulnerability assessment – topological analysis

While the mentioned literature suffers from a lack of data to apply statistical analysis, there are many studies which apply the concept of conditional vulnerability to evaluate a critical transport network, which allows them to build an assessment model without considering the probability of disruptive events. Consequently, this approach fits with topological analysis perfectly. Mattsson and Jenelius (2015) indicate that two traditional ways are topological vulnerability analysis of transport networks and system-based vulnerability analysis of transport networks, both of which allow researchers to assess the vulnerability in a quantitative way and also to avoid considering the probability of disruption.

There are many studies that successfully estimate the vulnerability of a transport network by topological analysis. Pagani et al. (2019) mapped the resilience and robustness of a public transport system network by considering peak hour and different types of railway systems and conducted topological analysis along with the concept of food webs. Zhang et al. (2018) assessed the resilience of metro networks and optimised recovery strategy by calculating how a node affects the connectivity and passenger flow in a network. The impact caused by removing a node is deemed as the vulnerability of the network as a result. Sun et al. (2018) applied the date of automated fare collection to analyse the cascading failures considering loading and redistribution. By considering the degree, betweenness and strength of nodes, the weak points in a network are recognised. Lin et al. (2020) established factors affecting passengers' behaviours under unplanned service disruption by building a nested logit model using stated preference data in Guangzhou, China and the result showed passengers have a higher possibility of mode shift when the disruption duration is long, occurs at peak hours and there are no alternative routes in the metro system. The study demonstrated what kinds of factors affect the decision-making of individual passengers during a disruption.

However, traditional topological analysis usually estimates the vulnerability of a network by calculating the connectivity of the affected origin–destination(O-D) pairs of a network after one or more nodes are removed from the network. First, it is commonly accepted that topological analysis applied in transport systems is not allowed to consider the behavioural response to disruption. Distinctly, we cannot consider how an individual passenger in the network reacts to the disruptive event, such as to what extent they reroute their trip and change their trip (Bešinović, 2020; Jenelius et al., 2006; Mattsson & Jenelius, 2015; Zhou et al., 2019). Additionally, topological analysis inevitably assumes not only the severity and frequency of disruptions are equal, but also that the condition of each node (or station when referring to a railway network) is the same. In other words, it only considers the result of the removal of each node and link regardless of its characteristic. Nevertheless, the type of disruption varies depending on the type of station (underground, ground or elevated station) and the time of year. For example, the exposure of trespass in a ground station is relatively higher than in an underground station. The exposure of leaf fall in autumn is relatively higher than in other seasons. This implies that the vulnerability of a network would depend on various attributes, which also implies the probability of disruptive events might be totally different from each other.

2.2.4 Vulnerability assessment – simulation and optimisation

In the context of either vulnerability or resilience of a complex system, some research looks at this issue from a numerical perspective. To simplify the complexity of a system, studies in this field aim to turn the performance of resilience into several metrics for further analysis. This approach to evaluation can be split into four types: data-driven models, topological models, simulation models, and optimisation models (Bešinović, 2020). Sometimes the probability theory models and fuzzy logic models are discussed in this context (Zhou et al., 2019); however, both approaches perform well on disassembling the sequence of incidents rather than ranking the performance of a system in terms of the resilience. Although all approaches mentioned above are quantitative methods, the four approaches described by Bešinović (2020) and the two approaches described by Zhou et al. (2019) have a completely different theoretical foundation. The first four approaches use reliability theory to interpret the concept of resilience, whereas the last two use probability theory to gain an understanding of how an accident happens. Having introduced the idea of data-driven models and topological models, simulation models and optimisation models are described in the remainder of this section.

First, simulation models are commonly applied in the analysis of large-scale transportation networks (Zhou et al., 2019). Due to the large-scale application, simulation models tend to use graph theory and modify the question in a topological way, which overcomes some constraints of a pure topological approach, such as the issue of operations dynamics (Bešinović, 2020). However, due to the complexity of simulation analysis, the literature only concentrates on a limited number of disruption analyses in a single study. The strength of simulation models is that it allows meaningful feedback in a given situation or strategies (Osei-Asamoah & Lownes, 2014). In the context of road transport, large-scale simulation can allow decision-makers to assess the link criticality of road networks (Kim & Yeo, 2017). However, the simulation is not only time-consuming, but also has massive data requirements (Bešinović, 2020), which makes it difficult to conduct and apply in practice.

On the other hand, the optimisation approach in assessing resilience derives the metrics to evaluate the resilience of either network-wide and scenario-specific cases by mathematical optimisation models (Bešinović, 2020). The objective of optimisation models in transportation research is mostly to address questions from the demand side or supply side. For questions from the demand side, it mainly concentrates on the choice of the path made by the road users or passengers. For instance, Szymula and Bešinović (2020) built an optimisation model to address the vulnerability of railway networks by considering both stranded and routed passengers; Sun et al. (2018) estimated the risk of stations in Beijing's rail network during a disruption by considering the redistribution of passenger flow, and Hong et al. (2017) considered the different level of passenger intermodal transfer distance preference between bus systems and subway systems to build a model to estimate the vulnerability of the public transport network. Studies which consider the demand side in transportation research are hard to conduct due to the difficulty of acquiring passengers' preferences. In contrast, there are many studies that aim to estimate the vulnerability of the system from the supply side. Researchers in this context divide a disruption into three sections: preparedness, response, and recovery (Zhou et al., 2019). The trade-offs between performance and resource allocation are widely discussed, for example, to optimise the benefit of resilience engineering in preparedness (Berle et al., 2011; Lämmel et al., 2010; Liao et al., 2018), to minimise the cost during the disruption or during the response (Jin et al., 2014; Marzuoli et al., 2016), and to estimate the performance of decisions in the recovery period following a disaster life cycle (Eid & El-adaway, 2017; Kaviani et al., 2017; Zhang & Miller-Hooks, 2015). The popularly used estimators (or dependent variables) in the context of rail and road transport from the demand side and supply side are shown in Table 2-2.

Table 2-2: Estimators of performance in transport networks considered in previous studies

	Estimators (in context of rail transport)	Estimators (in context of road transport)
Demand side	• Number of (un)satisfied demands • Total delay of passenger (mins) • Total welfare (dollar)	• Number of (un)satisfied demands • Total delay of passengers (mins) • Increased distance (km) • Total welfare (dollar)
Supply side (perspectives of operators)	• Remaining transport capacity (% of volume) • Cancelled/long-delayed transport services (no. of trains) • Recovery time (mins/hours/days) • Change in travel time (mins) • Frequency and duration of disruption* • Number of deaths or injuries • Service frequency (frequency-based service) (sec/mins)	• Average shortest paths (km) • Backup capacity (PCU) • Change in travel time (mins) • Frequency and duration of disruption* • Passenger load (%) • Passenger delay (mins) • Number of deaths or injuries • Recovery time (mins/hours/days) • Traffic flow (PCU/hour)
Supply side (perspectives of managers)	• Revenue vehicle-km (dollar) • Number of deaths or injuries • Recovery time (mins/hours/days) • Yearly disruption duration (mins)	• Repair/flow cost (dollar) • Recovery time (mins/hours/days) • Number of deaths or injuries

*can be once per day/month/year, depending on type of disruptive event

Both simulation and optimisation approaches can provide sufficient numerical results for decision-makers to do a further cost-benefit analysis. However, they cannot provide insights into the probability of events. The shortage limitation might result from the use of reliability theory in this context. Although the outcome of a disruption can be perfectly understood and estimated, the real risk is still unknown without assessing the probability of the disruption.

2.2.5 Vulnerability assessment – risk analysis

Compared with the concept of *conditional vulnerability*, *risk analysis* puts more emphasis on what could happen in the future. Literature usually associates risk with the answers to the following questions: what can go wrong?, what is the probability of that occurring?, and what are the consequences? (Kaplan & Garrick, 1981; Lemos, 2020; Rausand, 2013). A risk analysis is the application to provide the answer to these questions. In this section, the words of risk analysis are first introduced using an actual incident as an example (Section 2.2.5.1). The approaches of risk analysis adopted are then discussed(Section 2.2.5.2), and the association with vulnerability is also demonstrated (Section 2.2.5.3).

2.2.5.1 The words of risk analysis: an example

To clearly understand the process of risk analysis, the words used in this context should be first defined. A railway accident scenario of a signal passed at stop and near miss at Deansgate-Castlefield tram stop in Manchester (RAIB, 2020) is used to define each term in risk analysis. Definitions in the following contents are revised from the work of Rausand (2013).

This incident is summarised by RAIB (2020) as the following description:

"*At around 17:19 hrs on 17 May 2019, a tram passed through the centre platform of Deansgate-Castlefield tram stop on the Manchester Metrolink system, without making its* *scheduled stop. The tram exited the platform at around 9 mph (14 km/h) and then passed a stop signal. This placed it in the path of a second tram, which was approaching a junction as part of a signalled movement. The driver of the second tram saw the first tram approaching and was able to stop in time to avoid a collision.*" (RAIB, 2020)

First, many events are mentioned in the report. An *event* is generally considered as the component which occurs in a particular place during a particular interval of time in a sequence. For instance, "*a tram passed through*", "*tram stop on*" and "*without making its scheduled stop*" can be defined as an event. Among all events in a railway accident scenario, an *initial event* refers to the beginning of the scenario, whereas an *end event* refers to the end of the scenario. In this case, "*a tram passed through the centre platform of Deansgate-Castlefield*" is the initial event, and "*was able to stop in time*" is the end event. Another popular word applied to describe an initial event is a *hazardous event*. However, every event which is suggested as the potential cause of subsequent events which might damage the system can be a hazardous event. For example, regardless of the initial event mentioned above in this case, "*approaching a junction as part of a signalled movement*" can also be referred to as a hazardous event.

Second, to illustrate the accident scenario and the events more distinctly, the term "*hazard*" is commonly used to outline the source or condition resulting in harm *per se* or with a particular combination. A hazard can be property, interface, system, material or even procedure. It would damage nothing until out of control or triggered by another event. In this case, hazards can be "*a tram*", "*junction*" or "*the driver*". Regardless of hazards, *enabling events and conditions* are the events that enable hazards to become potentially harmful. In other words, the scenario can be put closer to the end event through the combination of *enabling events and conditions* and *hazards.* For example, "*without making its scheduled stop*" might be an enabling event.

However, only identifying hazards and enabling events and conditions is not enough to gain insight into an accident. Hence, active failures and latent conditions are commonly analysed in previous studies (Baysari et al., 2008; Madigan et al., 2016; Morais et al., 2020; Wiegmann & Shappell, 2017; Zhou & Lei, 2018). *Active failures* refer to the event(s) which trigger an unplanned event, and *latent conditions* refer to the hazardous components which have been existing but not yet contributed to an accident. In this example, the active failure could be the unawareness of the driver, and the latent conditions might be the poor design of the signalling system (which could have stopped the first tram automatically).

Third, even though there is no actual collision in this case, it could have happened due to the unusual action of the first tram. Therefore, such an incident is called a *near miss*, which means the event sequence could have developed until it ends with an accident, but it did not. In terms of *incident*, the literature concludes there are two meanings: the first one is an alternative word to near miss; another one is events which occurred in the past. In contrast, an *accident* means the sequence of unwanted events really results in damage to the tangible asset. For instance, if the second driver did not see the first tram and crashed into it, the outcome might be derailment, injury or even fatality, turning this event from incident or near miss into an accident. "*Consequence probability*" has also been proposed to refer to the probability of the near miss turning into an accident from a statistical point of view (An et al., 2011).

Last, the risk analysis would consider the *consequence* of an incident. Each hazardous event will result in a series of potential consequences associated with their probabilities. The effects of consequences could be fatality, injury and damage of property. Some studies even consider other long-term impacts, such as environment (Jones, 2001), society (Dreyer et al., 2010) and economic (Rosoff & von Winterfeldt, 2007). Generally, the *severity* is used to describe the seriousness of the consequences regardless of their measurements. In this case, if the collision really happens and results in derailment and injury, then “derailment” and “injury” are the sequences. On the other hand, the cost of repairing, healing and recovering the system is the severity of the consequence.

The words used in risk analysis example discussed above are summarised in Table 2-3.

Table 2-3: Summary of words used in risk analysis as applied to RAIB (2020), with definitions revised from Rausand (2013)

Terms	**Definitions**	**Examples**
An accident scenario	A sequence of events beginning with initial event and ending up with undesired consequence	signal passed at stop and near miss, Deansgate-Castlefield tram stop, Manchester
Event	The component which occurs in a particular place during the particular interval of time in a sequence	"*a tram passed through…*", "*tram stop on…*"
Initial event	The beginning of the accident scenario	"*a tram passed through the centre platform of Deansgate-Castlefield*"
End event	The end of the accident scenario	"*…was able to stop in time*"
Hazardous event	Every event which is suggested as the potential cause of events	"*…approaching a junction as part of a signalled movement*"
Hazard	The source or condition which can result in harm *per se* or with certain combinations	Property, interface, system, material or procedure
Enabling events and conditions	Event that enables hazards to become potentially harmful	"w*ithout making its scheduled stop*"
Active Failures	Event(s) which triggers unplanned event	The unawareness of the driver
Latent Conditions	The hazardous components which have been existing but not yet contributed to an accident	Poor design of signalling system
Accidents	The sequence of unwanted events really results in damage to tangible assets	If the second tram really crashes into the other tram

Table 2-3: Summary of words used in risk analysis as applied to RAIB (2020), with definitions revised from Rausand (2013) (continued)

Terms	**Definitions**	**Examples**
Incidents	1. The same definition as "near miss" 2. Events which occurred in the past	*"…was able to stop in time to avoid a collision*"
Near miss	The event sequence can have developed until it ends with an accident, but it does not	See example of "incidents"
Consequence probability	The probability of the near miss turning into an accident	The incident has 80% chance to become a collision
Consequence	The damage made by a hazardous event	Derailment and injury
Severity	The seriousness of the consequences regardless of their measurements	Cost of derailment and injury

2.2.5.2 Methods for risk analysis

Researchers in risk analysis have devoted much work to mitigate the impact of accidents. One popular approach is to reveal the causes of previous accidents in the case they repeat themselves, also known as sequential accident models (Rausand, 2013). Hence, many models have been proposed for the purpose of understanding the causality of accidents. One of the first models in the risk analysis field was Heinrich's domino theory (Heinrich, 1941), which argues that an accident consists of a series of events and a linear one-to-one progression would eventually lead to an unexpected result (Kim & Yoon, 2013; Rausand, 2013). Subsequently, Heinrich's domino was revised by Bird and Germain (1986) and Rasmussen and Svedung (2000) and some critical components have been added such as the concept of property loss (Bird & Germain, 1986) and hazardous environment (Rasmussen & Svedung, 2000).

A considerable number of requests have been made to construct the sequential model in a system-based way (Reason, 1990, 2000; Rosoff & von Winterfeldt, 2007), leading to many accident analysis models. Of these models, it has been widely agreed that Reason's Swiss cheese model has been very influential (Kim & Yoon, 2013; Zhou & Lei, 2018). The Swiss cheese model proposes that an accident is triggered by the combination of a series of active failures and specific latent conditions (Reason, 1990). Reason (1990) described the barriers as the cheese slices, and the active failure or latent condition as the hole in a slice. Once all barriers are penetrated at the same time (holes), the unexpected accident would happen (Reason, 1990). This model first breaks down the isolated event into active factors and latent conditions, which allows decision-makers to learn from previous accidents and control risk by identifying hazardous conditions beforehand.

Nowadays, the complexity of railway systems has increased dramatically due to the development of technology. Thus, several complex models have been designed to explore accidents from different levels and perspectives including man–technology–organisation (MTO) analysis for analysing the importance of humans, technology and organisation within an accident (Sklet, 2002), the management oversight and risk tree (MORT) approach for analysing an accident in a deductive way (Johnson, 1973), and the software–hardware–environment–liveware (SHELL) model for clarifying the relationship between the environment, infrastructure and human factors (Wiegmann & Shappell, 2017).

Despite many critical factors that have been identified in the literature, such as human error (Baysari et al., 2008; Morais et al., 2020; Wiegmann & Shappell, 2017), organisational failure (Shappell & Wiegmann, 2000), infrastructure failure (Little, 2002; Restrepo et al., 2009) and environment hazards (Dawson et al., 2016; Hong et al., 2015), only limited studies have concentrated on the role *interface* plays in an accident. It has been proven that interface can significantly influence the process safety management system (Kelly & Berger, 2006), regardless of the physical or organisational interface. For physical interface, the importance of the people, property and protection interfaces of urban underground railway infrastructure and its environment at different levels of consideration has been revealed (Darroch et al., 2016, 2018). For organisational interfaces, the issue of organisation interface management in the context of a construction project (Li & Guo, 2011; Parraguez et al., 2016) or large project, such as projects in the petroleum industry (Milch &Laumann, 2018, 2019; Ventroux et al., 2017), have been widely discussed. Additionally, there is a small group of studies that aggregate empirical studies and build theory (Jiang & Kong, 2013; Johnsen et al., 2006). Existing studies on organisation interface management and suggested best practice for improving safety at cultural interfaces proposed by Johnsen et al. (2006) are shown in Table 2-4 and Table 2-5 respectively.

Among all studies relating to interface management, only a few of them see the topic of organisational interfaces from the perspective of risk management. Although some studies have tried to manage the risk raised by the failure of interfaces within the organisation or between organisation and environment (Cedergren, 2013; Pires & Mosleh, 2011), the causality of the interface failure has never been clearly explored.

Table 2-4: Literature on organisation interface management

Object	Data resource	Methodology	Result	Resource
Dealing with urban space issues from both engineering and legal perspectives	Interviews and primary and secondary sources	A case study – Glasgow subway	A conceptual framework for describing three principal interfaces identified as presence, property and protection.	(Darroch et al., 2016)
Exploring interface between underground urban transport infrastructure and its environment and how they could and should be managed effectively	London Underground Ltd. Plan	A case study – A building over the metropolitan line north of Farringdon station	Understanding the presence, property, and protection interfaces of urban underground infrastructure and its environment at different levels of consideration.	(Darroch et al., 2018)
Proposing a process-based approach for interface management of mega capital projects	N/A	Interface management	Steps for building Interface Management System have been developed.	(Shokri et al., 2012)
Identifying the critical "interfaces" between the many participants in a process safety management system	N/A	Interface management	Basic steps to analyse critical interfaces and implement measures to manage them have been developed.	(Kelly & Berger, 2006)
Construct the comprehensive evaluation index system of organisation interface management of construction project	Experts consultation	AHP framework	Developing an AHP framework which contains 25 quantification evaluation indexes.	(Li & Guo, 2011)

Table 2-4: Literature on organisation interface management (continued)

Object	Data resource	Methodology	Result	Resource
Propose a new approach that characterises process interfaces as organisation networks	A biomass power plant project	Clustering analysis	Revealing a relationship between the structure and composition of the process interfaces and reported interface problems.	(Parraguez et al., 2016)
Gain better understanding on interface management theory and synergetic theory	Literature review		Proposing an operation Model of Collaborative Management of Organisation Interface of Large-scale Projects.	(Jiang & Kong, 2013)
Propose a new approach to improve collaboration between interdependent actors (oil and gas projects)	Pazflor project	Actors/ Design Structure Matrix	Identifying the vulnerability of the collaboration between actors and the detection of complex phenomena.	(Ventroux et al., 2017)
Revealing challenges related to implementation of recommendations from accident investigations	Swedish railway sector	Interview/case study	Two factors are found: 1. A trade-off between being insider and outsider to the industry. 2. A trade-off between micro-level and macro-level factors.	(Cedergren, 2013)
Exploring how interorganisational complexity is managed on a petroleum-producing installation	Norwegian petroleum industry	Semi-structured interviews	Interorganisational challenges are identified: 1. Coordinating work processes among companies. 2. Variations in experience among sharp-end workers from sub-contractor companies.	(Milch & Laumann, 2018)

Table 2-4: Literature on organisation interface management (continued)

Object	**Data resource**	**Methodology**	**Result**	**Resource**
Improving safety at cultural interfaces	International Union of Railways (UIC)	Questionnaires/ exploration of scenarios/ STEP diagram	See Table 2-5. Suggested best practice for improving safety at cultural interfaces proposed by Johnsen et al. (2006)	(Johnsen et al., 2006)
Organisational interface failures: A historical perspective and risk analysis framework	N/A	N/A	Three failure categories in terms of interface: 1. Communication interfaces 2. Coordination interfaces 3. Collaboration interfaces	(Pires & Mosleh, 2011)
Investigating petroleum incidents considering the interfaces between companies	22 reports from Norwegian Continental Shelf	Braun and Clarke's version of thematic analysis	Four themes were identified: 1. Ambiguities in roles and responsibilities between personnel from different companies. 2. Inadequate processes to ensure sufficient competence across interfaces. 3. Inadequate quality control routines across organisational interfaces. 4. Communication breakdowns between companies.	(Milch & Laumann, 2019)

Table 2-5: Suggested best practice for improving safety at cultural interfaces proposed by Johnsen et al. (2006)

Key elements	Definition / Example
No tolerance of “Grey areas” of responsibility.	To have clarity in task definitions and responsibilities across interfaces.
Obligation to report any condition that could imply a risk for other companies.	To share databases regarding safety events and the resulting recommendations among all parties.
The use of protocols or formalised communication templates.	To reduce difficulties in understanding through pre-determined protocols and forms.
Harmonisation of procedures by project teams across organisational boundaries.	Groups should meet face-to-face, to establish standard procedures and create confidence and common understanding.
Common rules and procedures.	To decide on one set of rules and change this as little as possible.
Intensive standardised training for operators, focusing on communication and handling of deviations.	To establish standard mental models and understanding that can be shared among the operators.
A similar model for identifying and managing risks and the resources to control risks.	To systematically address most difficult issues in the conceptualisation of risk management.
Sharing experience	To share experiences to provide an opportunity to learn from each other.

2.2.5.3 Risk analysis in the context of the railway industry

Risk analysis in the context of the railway industry usually concentrates on either the role latent conditions play in an accident (Elms, 2001; Leveson, 2004; Madigan et al., 2016), or how an accident occurs in terms of active failure (Baysari et al., 2008; Elms, 2001; Kim & Yoon, 2013; Zhou & Lei, 2018). The aim of the former is to focus on how to build a model to explore the importance of latent factors in an accident, whereas the latter aims to emphasise the sequential accident model, which can help to integrate knowledge from historical incidents and gain an understanding on how to prevent accidents in the future.

For the model which emphasises latent factors, the human factors analysis and classification system (HFACS) is widely applied in this context (Baysari et al., 2008; Raslear, 2006; Yu et al., 2018; Zhou & Lei, 2018). HFACS is a well-known model derived from Reason’s Swiss cheese model and man-made disasters theory (Turner & Pidgeon, 1997). HFACS was first introduced to analyse aviation accidents, which only highlights specific active failures and latent conditions (Shappell &Wiegmann, 2000). However, due to the similarity between the aviation and railway industries in terms of causes of accidents, HFACS has also been popularly applied in railway safety (Baysari et al., 2008; Kim & Yoon, 2013; Zhou & Lei, 2018). Another popular model is systems theoretic accident model and process (STAMP), which argues that the cause of an accident is not events, but a lack of constraints. The risk can be well controlled once the chain of constraints in an organisation is appropriately managed (Leveson, 2004). STAMP is popularly used in case studies, such as the China–Jiaoji railway accident (Ouyang et al., 2010), the China–Yongwen railway accident (Song et al., 2012) and high-speed railway accidents (Liu et al., 2015).

On the other hand, studies in the field of sequential accident models either apply frameworks built on the basis of complicated models (Baysari et al., 2008; Madigan et al., 2016; Zhou & Lei, 2018), or purely conduct graphical models to demonstrate the causality between factors, such as Bayesian networks (Castillo et al., 2016; Kim & Yoon, 2013) and fault tree analysis (Huang, Liu, et al., 2020; P. Liu et al., 2015). However, due to the difficulty of acquiring statistical data, most studies in this field derive the model based on only one or a few cases, which has been criticised for lack of validity due to insufficient data (Kim & Yoon, 2013; Zhou & Lei, 2018). Despite these criticisms, proponents stress the importance of gaining the probability of each event within an accident by enlarging the scale of the database and building the model by highlighting the critical elements in each accident manually. For instance, Kim and Yoon (2013) conducted 80 rail accident investigation reports from the UK and extended the HFACS model shown as Figure 2-3; and Zhou and Lei (2018) analysed 407 railway accidents or incidents in China. Existing studies in this context are integrated in Table 2-6.

Although it seems that previous studies have successfully analysed accident reports manually, which is large enough for further statistical analysis, such a labour-intensive analysis method has limitations. For instance, for a cross-country analysis, the volume of data invlovded would be extremely huge, and the cost associated with retrieving and processing such data would be unaffordable. Additionally, to propose and test a model with new variables they need to be abstracted from reports, all the data needs to be read again. The reason for the difficulty of obtaining data in an efficient way may be due to a lack of knowledge on how to apply state-of-the-art technology to help acquire data, which in turn, can result in a number of gaps in the existing literature. This is explored in this thesis and details of the technology used are introduced in the following chapter.

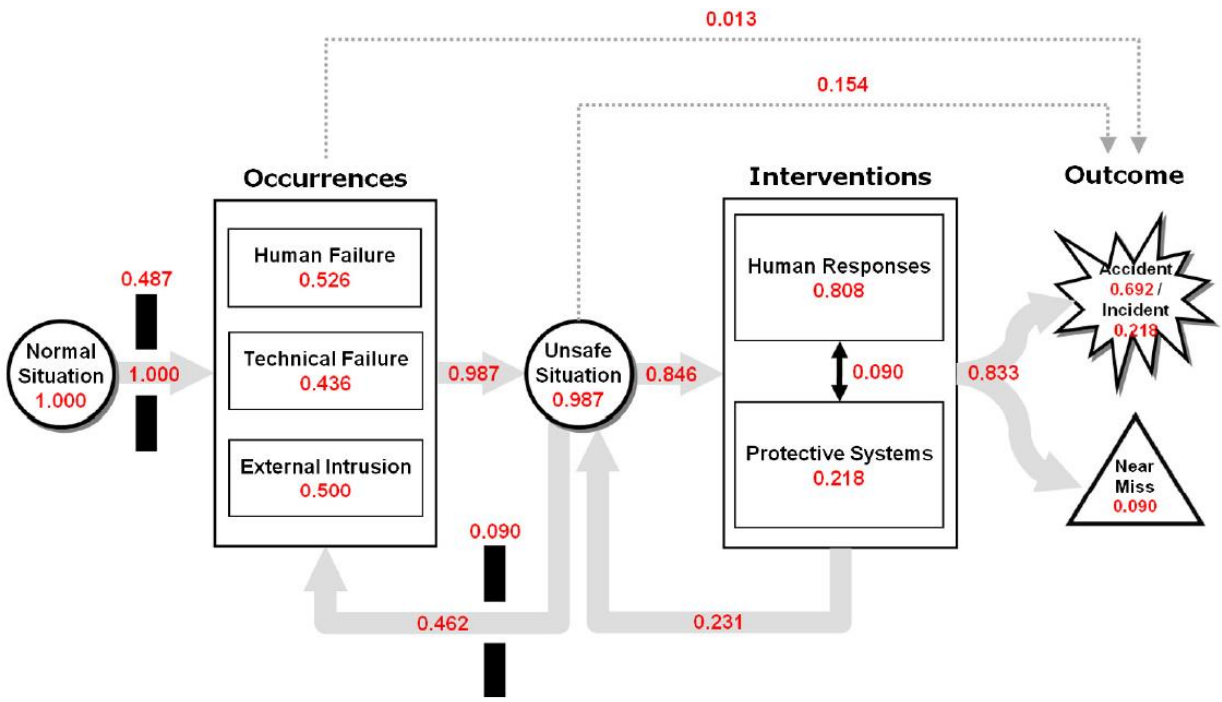


Figure 2-3: The framework and occurrence rate of incident components proposed by Kim and Yoon (2013)

Table 2-6: Existing studies in the context of railway risk analysis

<table>
<tr><th>Object</th><th>Data resource</th><th>Methodology</th><th>Result</th><th>Source</th></tr>
<tr><td>Understand the relationship between active and latent factors</td><td>78 incident reports
5 types of incident
(UK)</td><td>Expert judgment (raters)
HFACS</td><td>Highlight latent factors at the supervisory and organisational levels.
Suggest a new factor – operational environment.</td><td>(Madigan et al., 2016)</td></tr>
<tr><td>Identifying human errors associated with rail accidents/incidents in Australia</td><td>40 rail safety investigation reports (Australian Transport Safety Bureau, Office of Transport Safety Investigations, Victorian Department of Infrastructure, Queensland Transport)</td><td>Expert judgment (raters)
HFACS</td><td>Nearly half the incidents resulted from an equipment failure.
Most of these were the product of inadequate maintenance or monitoring programs.</td><td>(Baysari et al., 2008)</td></tr>
<tr><td>Summarising the nature of the rail safety management problems</td><td colspan="3">Dimension of rail safety problems
Focus on management, be distributed, be differentiable, be heavily dependent on human factors, have both internal and external consequences, be reasonably simple, be transparent, be subject to external scrutiny</td><td>(Elms, 2001)</td></tr>
<tr><td>Presenting a new accident model based on systems theory concepts</td><td>Model is derived from control theory</td><td colspan="2">The cause of an accident is not events, but a lack of constraints. Systems Theoretic Accident Model and Process (STAMP) is built based on control loops and process models.</td><td>(Leveson, 2004)</td></tr>
</table>

Table 2-6: Existing studies in the context of railway risk analysis

Object	**Data resource**	**Methodology**	**Result**	**Source**
Propose an accident causation model for the railway industry	80 accident reports from RAIB (UK)	Manually reconstruct event sequences and calculate the occurrence rate	Shown in Figure 2-3	(Kim & Yoon, 2013)
Exploring paths between categories in HFACS	407 railway accident/incident reports in China	Expert judgment (raters) HFACS	Using statistical analysis to determine significant associations in HFACS	(Zhou & Lei, 2018)

2.3 Analysis of railway accident reports

According to the literature reviewed in previous sections, the outcome of each incident can be easily estimated through numerical analysis, such as optimisation, simulation and topological approaches. However, the probability of each component in an accident is still not fully understood. Although literature has tried to apply probability theory to address this issue, it suffers from the lack of analysable statistical data. Conversely, there are sufficient data items in textual format. In the context of railway safety in practice, each incident has been recorded in detail by the independent investigator from its jurisdiction. The documentation and the writing style are different among different jurisdictions. For instance, the framework of railway accident reports would be different between investigators from the US and the UK. Additionally, the English language used in different jurisdictions is not quite the same, which increases the difficulty of doing textual analysis. To address the issues abovementioned, this section introduces the characteristics of railway accident reports by independent investigators from various jurisdictions. The issue of doing textual analysis and the linguistic problem is discussed in Chapter 3.

2.3.1 The jurisdictions of the railway industry

Once a railway accident occurs the relevant accident investigation body begins to investigate. The aims of the investigator will vary according to jurisdiction and are foundationally different and can be broken down into two categories based on their purpose: the first one is to conduct the investigation to improve railway safety and to prevent further accidents; the other one is to apportion blame and liability. In this section, the jurisdictions of railway accident investigators in the United Kingdom, the United States, Australia, and Canada are introduced. These four countries all have a sound documentation system of incident reporting and several railway accident investigation organisations, which are worth further research.

2.3.1.1 United Kingdom

There are several railway accident investigators in the United Kingdom with different jurisdictions, including the Rail Accident Investigation Branch (RAIB), the Crown Office and Procurator Fiscal Service (COPFS) in Scotland, the police and other safety authorities.

The investigation of railway accidents is conducted by either the local railway safety regulator or the railway company. After the Ladbroke Grove rail crash in October 1999, a requirement for an independent rail accident investigation unit was raised by the public. Subsequently, RAIB was established to independently investigate accidents causing death, serious injuries or extensive damage. Additionally, any non-serious railway accident or incident is also eligible for RAIB to conduct an investigation (RAIB, 2019). The establishment of RAIB has made a clear distinction between the prosecution-oriented and safety development-oriented investigator. RAIB offers the technical investigation into the causality and consequence of an accident (RAIB, 2005). The outcomes of the RAIB investigation usually turn into the critical input of prosecuting bodies' investigations or the basis of further regulation (RAIB, 2005). Additionally, the authority whom recommendations are addressed to is required to provide sufficient action on the recommendations made by RAIB.

On the other hand, the purpose of other investigators like Scotland's COPFS, the police and other safety authorities is neither to apportion blame and liability or bring about prosecution if there has been a breach of the law. Most of them are prosecuting bodes, which might result in a report overemphasising the allocation of responsibility. For instance, procurators fiscal and the police in Scotland conduct criminal and death investigations as appropriate, which usually turn into a criminal prosecution and a fatal accident report.

2.3.1.2 United States

The National Transportation Safety Board (NTSB) has independent jurisdiction over investigating railway accidents at the national level. Although NTSB mainly emphasises the investigations of civil aviation accidents, it is also in charge of major accidents in other modes of transportation (NTSB, 2016). As with the independent railway accident investigators in other countries, NTSB is not responsible for any regulatory agency. The only purpose of NTSB is to investigate the railway accident solely, and offer appropriate recommendations, which could help railway operators or regulators to improve safety.

However, unlike the clear jurisdiction over the investigation of civil aviation accidents in the United States (NTSB, 2014a), the responsibility of NTSB in a railway context is ambiguous. NTSB claims that it would investigate “Any accident that occurs in connection with the transportation of people or property that, in the judgment of the NTSB, is catastrophic, involves problems of a recurring nature or would otherwise carry out the intent of its authorizing statutes” (Lauby, 2016; NTSB, 2016), and NTSB also proposes to have primary jurisdiction over railway accidents or incidents resulting in death or severe property damage (NTSB, 2016). In other words, accidents which NTSB refuses to investigate would be taken over by either the local safety authorities, or the United States Federal Railroad Administration (FRA), which is charged with promulgating, developing and enforcing rail safety regulations (FRA, 2010).

In some cases, multiple jurisdictions would conduct their own investigation into the same incident separately. Therefore, the way each investigator analyses the accident and the conclusion made by different investigators would depend on their jurisdictions. For instance, there are two accident investigation reports for the CSX freight transportation derailment and fire incident in August 2017, conducted by NTSB (NTSB, 2020) and FRA (FRA, 2020). However, this case covers the issue of not only derailment, but also weather alerts and hazardous materials. In terms of the conclusions, FRA simply suggested that defective equipment did not lead to the cause or severity of the accident (FRA, 2020), whereas NTSB claimed it is still undertaking the investigation and will make recommendations to FRA, the Pipeline and Hazardous Materials Safety Administration, CSX Transportation, the Brotherhood of Locomotive Engineers and Trainmen, the International Association of Sheet Metal, Air, Rail and Transportation Workers, and Trinity Industries (NTSB, 2020).

2.3.1.3 Australia

The investigation of railway accidents is selective, depending on the interests of jurisdictions. The Australian Transport Safety Bureau (ATSB) plays the role of improving railway safety by investigating railway accidents independently at a national level and formulating appropriate recommendations. As with other national railway accident investigators in other countries introduced above, the ATSB aims to conduct “no-blame” investigations according to a breach of the law (ATSB, 2009; Transport Safety Investigation Act 2003 (Cwlth), 2016). Instead, ATSB concentrates on how to enrich the understanding of accidents and safety management and leaves apportioning blame or liability to other safety authorities.

However, ATSB does not investigate all incidents occurring in railway systems. Only accidents which can provide further insight are considered to be investigated. For instance, if one accident has been identified as a repetitive event in nature, it would not be investigated. Another case is that once early evidence has indicated the accident is caused by a criminal act, then ATSB would leave the jurisdiction over the investigation to the police (ATSB, 2009).

Despite ATSB having no legislative power to make recommendations into compulsory implementations, ATSB is still eligible to require a detailed response toward the implementation of recommendations and to disclose to the public (ATSB, 2009).

In addition, there are other safety authorities in Australia which conduct railway accident investigations, such as state and territory rail regulators, policy makers, and railway operators. However, most of them investigate to bring about a prosecution intentionally, which means their reports mainly discuss the apportionment of liability.

2.3.1.4 Canada

The Transportation Safety Board of Canada (TSB), established in 1990, is the agency of the Government of Canada with full jurisdiction over investigating aviation, rail, marine and pipeline transportation. According to TSB, except for the Department of National Défense and the Royal Canadian Mounted Police, it is the only authority that has jurisdiction over conducting transportation accident investigation to find the causes and contributing factors of the accident (TSB, 2014a). This indicates TSB can be considered as the only organisation which is able to enhance transportation safety by investigating accidents. The reporting system and documentation are well-designed, and the TSB reports are suitable for conducting further analysis, such as text analysis. The recommendations TSB makes are not about blame.

2.3.1.5 Other jurisdictions

Apart from the jurisdictions mentioned above, there are several railway accident investigators in other countries. Some of them not only have jurisdiction over investigating transportation accidents but have also been authorised to conduct investigations without considering allocating liability. For instance, the Dutch Safety Board (DSB) in the Netherlands, formed in 2005, has jurisdiction over the various transport sectors and in the fields of defence, industry and trade, healthcare and more; the Taiwan Transportation Safety Board (TTSB) in Taiwan, formed in 1998, is an independent government agency which is responsible for investigating major transportation accidents in aviation, railways, waterways and highways in Taiwan; the

Aviation and Railway Accident Investigation Board (ARAIB) in South Korea is independently in charge of investigating aviation and railway accidents; and the Japan Transport Safety Board (JTSB), formed in 2008, has full authority to investigate transportation accidents and increase the safety of transportation.

Table 2-7 compares jurisdictions and their objectives from different countries.

Table 2-7: Investigating jurisdictions and their objectives from four countries

Country	Jurisdiction	Objectives
United Kingdom	RAIB	Improving railway safety and preventing further accidents
	COPFS, Scotland	Apportioning blame and liability, bringing about a prosecution if there has been a breach of the law
	Other safety authorities	
United States	NTSB	Improving railway safety and preventing further accidents, especially for those with death or severe property damage
	FRA	Mixed objective with the power of regulation
	Other safety authorities	Apportioning blame and liability, bringing about a prosecution if there has been a breach of the law
Australia	ATSB	Improving railway safety and preventing further accidents
	Other safety authorities	Apportioning blame and liability, bringing about a prosecution if there has been a breach of the law
Canada	TSB	Improving railway safety and preventing further accidents
	Other safety authorities	Apportioning blame and liability, bringing about a prosecution if there has been a breach of the law

2.3.2 Principal content of railway accident reports made by jurisdictions

The writing style of railway accident reports from different investigators varies, and is based on the act made in its jurisdiction. Almost all investigators provide not only a preliminary report when the very first findings are available, but also a finalised report once the whole investigation is done. This section introduces the content of the railway accident report from the investigators. To reduce the complexity of the analysis, only finalised reports are considered.

2.3.2.1 RAIB, UK

RAIB conducts an investigation in accordance with the Railways (Accident Investigation and Reporting) Regulations 2005 (RAIB, 2005). The act also regulates the principal content of the report. The report first provides a summary of the investigation, which briefly outlines the descriptive facts such as the occurrence, synopsis and the consequences. The direct contributing factors, underlying causes, and main recommendations must also be summarised. The summary of the accident is explicitly disclosed, including the involved trains, people, organisations and facilities. Immediate facts of the occurrence are also revealed. Then, records of investigations and inquiries are provided as the evidence, including rules and regulations, man–machine–organisation interface, and the safety management system, etc. Finally, the analysis and conclusions are made, and recommendations are stated.

2.3.2.2 NTSB, US

The conduct of rail accident reports by the NTSB is under the Guidance on Style for NTSB Written Products (NTSB, 2014b). The role of documentation NTSB applies is not only accidents in the field of railways, but also other modes of transport, such as aviation, marine and pipeline transport. The body of a report is relatively rigorous compared with investigators in other jurisdictions, and has an executive summary, body of the report, factual information, analysis, conclusions and recommendations. The content under each heading is strictly regulated and the terminology must be consistent in the documentation. The contributing factors NTSB considers are comprehensive, including personnel information, train and mechanical information, operational information, management information, meteorological information and medical and pathological information (for more details, see NTSB, 2014b).

2.3.2.3 FRA, US

Compared with the investigation conducted by the NTSB, the structure of railroad accident reports by the FRA is simple. The content only consists of a synopsis (which has the same content as the summary in other jurisdictions), general information, sketches, narrative, analysis and conclusion. Moreover, FRA conducts a “FRA factual railroad accident report”, which only aims to recognise the probable cause and contributing factors rather than give recommendations. Interestingly, the information FRA provides mainly contains active factors like human errors and infrastructure failure. The people and equipment involved in an incident are detailed. On the contrary, latent conditions, such as safety culture or organisational management, are not mentioned in the reports.

2.3.2.4 ATSB, Australia

ATSB produces investigation and accident reports based on Railway Accident Investigation Guidelines for Railway Network Owners, Railway Operators, and Emergency Services Personnel (ATSB, 2009), which incorporates the Transport Safety Investigation Act 2003. The writing style of ATSB is similar to RAIB, and reports also contain a summary, the investigation, findings, conclusions and recommendations. Contributing factors are also revealed in the report, but only active factors are included. For instance, in the report of the derailment of grain train 8838N at Narwonah, New South Wales in October 2017, ATSB only concluded that poor track condition resulted in the derailment, and maintenance of defects in this section of track cannot successfully prevent the reoccurrence of defects (ATSB, 2020). However, the latent conditions behind this fact, such as the failure of organisational communication, are not revealed by ATSB, making the railway system unable to prevent such an accident efficiently.

Despite the lack of understanding on latent conditions, ATSB has comprehensive documentation system on the recommendations it makes. Once ATSB makes recommendations in an incident, they are saved in an online repository called “Rail safety issues and actions,” which contains all the recommendations from previous railway accident investigations conducted by ATSB previously. ATSB also provides audio of its reports on its website, so details of the accident reports can also be “listened” to.

2.3.2.5 TSB, Canada

Under the Canadian Transportation Accident Investigation and Safety Board Act (S.C. 1989, c. 3), the reports conducted by TSB do not have a specific writing style. Instead, the framework of accident report TSB uses depends on the scale of the accident. For instance, the difference of scale between an incident that destroyed 40 buildings and 53 vehicles (TSB, 2014b) and an incident of Main-track train derailment (TSB, 2019) is significant. On the other hand, because of the lack of regulated writing style, the content of reports also varies depending on the team conducting the investigation.

However, TSB has two remarkable elements in its documentation. First, TSB tracks whether the recommendations it has made are satisfied or not, and all information is put on the “Rail transportation safety recommendations” website (see TSB, 1990). TSB monitors the execution of each recommendation until it meets the criteria of Satisfactory. Second, TSB occasionally lists in its reports previous cases that are similar to the incident.

Table 2-8 summarises the contents of accident reports from different investigators.

Table 2-8: Summary of contents of accident reports from different investigators

	RAIB	**NTSB**	**FRA**	**ATSB**	**TSB**
Summary	√	√	√	√	√
Narrative	√	√	√	√	√
Detail/Result	√	√	√	√	√
Causality analysis	√	√	√	√	√
Sketches	Δ	Δ	√	Δ	Δ
Similar cases (if available)	√	X	X	X	√
Recommendations	√	√	X	√	√
Near miss	√	X	X	Δ	√
International comparisons	X	X	X	X	X

Note: √: always, Δ: limited, X: never.

2.3.3 Revealed and unrevealed factors in literature about the causes of railway accidents

Most studies apply a framework derived from other contexts to do an empirical analysis in the field of the railway industry. Hence, some prevalent factors which have been identified in other contexts have been widely accepted in the railway field, including human factors (Kim & Yoon, 2013; Zhou & Lei, 2018), technical or design failure (Kinnersley & Roelen, 2007; Yu et al., 2018), organisational issues (Clarke, 1998) and environmental impact (Hong et al., 2015). For human factors, a considerable number of studies have identified several critical factors and constructed many comprehensive models in terms of blame (Whittingham, 2004), reliability (DeFelice & Petrillo, 2011), and inappropriate performance (Kyriakidis et al., 2015). The human error has been considered to be a significant cause of most accidents within human factors (Woods & Cook, 2003). Recent research focuses on the interpretation of human errors from the perspective of Ergonomics and Human Factors (EHF). It concludes that there has been a shift from human performance to systems ergonomics and suggests the promotion of proactive system design incorporating human behaviours and reactions (Carayon et al., 2013; Read et al., 2021).

For technical or design failure, these factors are seldom discussed independently (Fan et al., 2015), but literature usually explores technical or design failure factors through exhaustive analysis (Kim & Yoon, 2013). Historically, the difficulty of classifying the content of railway accidents and gaining analysable statistical data means that the literature seldom conducts a comparison among all variables, given that the number of railway accident reports is sufficient to offer more insights. Nevertheless, the textual data in accident reports previously needed to be analysed manually, which limited the number of attributes considered in the literature.

Organisational issues are usually considered as part of human error, especially since the HFACS model deems organisational influences as one factor which would affect human behaviours (Baysari et al., 2008; Madigan et al., 2016; Zhou & Lei, 2018). Last, environmental factors have usually been considered as an independent research topic, and case study is popularly used to gain an understanding of different environmental impacts, such as floods (Hong et al., 2015) and sea-level rise (Dawson et al., 2016).

Regardless of many studies which have discovered elements which would trigger a railway accident individually, none of them have looked at railway accidents from the perspective of interface failure. Several safety-oriented industries have considered this issue as a critical factor, such as petroleum and aviation (see Table 2-4). However, in the railway safety literature reviewed, this issue has not been highlighted. Without an understanding of interfaces failure, decision-makers could be left in a void in terms of analysing railway accident reports because they only take individual factors into account instead of considering the elements caused by the interface of two or more factors.

Last, the recommendations made by each jurisdiction are seldom discussed in a railway safety context. Literature mainly concentrates on the role the investigator plays in a railway accident. For instance, it has been revealed that investigators have difficulty in determining the scope of recommendations and allocating responsibility for the tasks derived from recommendations (Cedergren, 2013). Additionally, previous studies have not only found that the reason an accident always repeats itself is the lack of systematically studying learning and understanding even though the same recommendations have been made every time after an accident (Drupsteen & Hasle, 2014; Wrigstad et al., 2014), but also emphasised the gap between the recommendations made by investigators or academia and implementation by industry (Brath, 2020; Underwood & Waterson, 2013).

2.3.4 Current applications for safety recommendations in the railway industry

Most railway bodies are jurisdiction-based industry, they are supervised by the railway authority in their jurisdictions. The rail investigation body is the branch of the railway authority, which aims to investigate accidents and make recommendations to enhance railway safety. The public body or authority to whom a recommendation is addressed in the jurisdiction must make appropriate reactions and provide the details of measures to secure implementation (RAIB, 2019). The reports and the status of the recommendations are openly published as well. Despite the increased awareness of the importance of how to efficiently implement recommendations, in the context of railway safety management can be seen to have ignored the issue of the efficiency of learning across jurisdictions due to several issues, such as language barriers and legal obligations. For instance, the recommendations by RAIB are not an official mandate for other regimes. Additionally, the lack of synchronisation in terms of accidents and recommendations among jurisdictions exposes the railway industry to risks. Such a knowledge gap increases the vulnerability of the railway industry and allows an accident more likely to re-occur.

In contrast, several international industries, such as civil aviation and the nuclear energy industry, have recognised this issue and funded international organisations to make a series of regulations for operators in the industry to prevent serious accidents. For instance, the

International Civil Aviation Organisation (ICAO) has been authorised to maintain an administrative and expert bureaucracy to standardise air transport policy. Although the ICAO claims that it is not an international aviation regulator, the Chicago Convention on International Civil Aviation and other countries spontaneously follow the Articles and Annexes provided by ICAO to ensure the best safety performance (ICAO, 1944).

Several international railway safety organisations, such as International Railway Safety Council (IRSC) and International Union of Railways (IUR), have members from many countries. IUR has even set up a standard of terminology and data, which is also known as the UIC leaflet. However, the power of these institutions is not significant because, unlike the aviation industry, the railway industry primarily follow regulations within its jurisdiction. Railway companies in each jurisdiction do not receive any punishment if they do not follow the guideline established by the international railway safety organisation. Hence, the connection between jurisdictions is poor, resulting in inefficient delivery of knowledge relating to railway safety, and the risk eventually emerges.

In conclusion, the lack of awareness of learning across jurisdictions in the railway arguably results in poor understanding of railway safety management in terms of knowledge delivery. An in-depth understanding of estimating the efficiency of sharing railway accident knowledge may be crucial to improving overall railway safety around the world. In order to overcome the difficulty of sharing knowledge, some researchers have considered applying textual data (reports) directly to gain horizontal insights through the state-of-the-art technique of Natural Language Processing (NLP).

2.3.5 Application of NLP in the context of railway safety

The development of NLP allows researchers to identify insights from accidents and disseminate the information efficiently. NLP is a technique that enables a computer to analyse textual data and generate a summary or horizontal conclusion through reading an enormous number of words in articles. The extracted information can be the input for further downstream analysis (more details are given in Chapter 3). NLP has been widely used in many contexts, such as finance (Xing et al., 2018; Yang et al., 2020; Yildirim et al., 2019), healthcare (Bayat et al., 2009; Khetan et al., 2020) and the chemical industry (Du et al., 2020; Luo et al., 2018; Zhang et al., 2016).

The studies of accident causality analysis in literature are usually qualitative analysis or limited case study (Ouyang et al., 2010; Santos-Reyes & Beard, 2006; Santos-Reyes & Beard, 2009), and quantitative analysis of accident causality analysis requires labour-intensive data setup (Kim & Yoon, 2013; Kyriakidis et al., 2015; Zhou & Lei, 2018). With the growing knowledge in the context of NLP, researchers turned to seek solutions from the field of computer science. In the early stage, several models were designed to classify accident reports into several categories, like human error, technology issues and organisation issues (Heidarysafa et al., 2019; Li et al., 2018; Syeda et al., 2017; Yu et al., 2018). The approaches of supervised and unsupervised learning have been applied on the document-level causal classification, and the conceptual model is demonstrated as Figure 2-4. Although dealing with the causality extraction problems from the document-level perspective is intuitive, actual rail accidents contain several dimensional failures instead of one single cause. Hence, such an approach would oversimplify the content of accident reports, and the natural causality between elements in a rail accident would not be revealed.

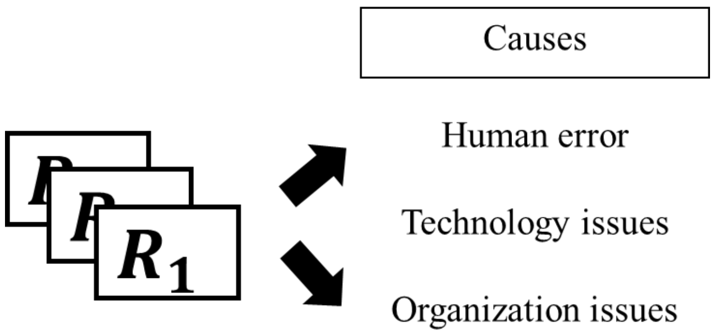


Figure 2-4: The concept of document-level causal classification

To understand the complex causality demonstrated in the accident reports, researchers began to switch the scale of analysis from document level to sentence level. Several studies designed models to identify the characteristics of the sentences and to extract “causes” and “results/effects” from chosen sentences in rail accident reports (Hua et al., 2019; Li et al., 2018). The conceptual framework of the models used is demonstrated in Figure 2-5. However, the performance of these models is limited. The main reason might be that actual accident reports contain a very complex structure of sentences, and important connections exist between sentences in the reports. If the model only considers causes and results rather than the connections between sentences, the causality still cannot be explained explicitly.

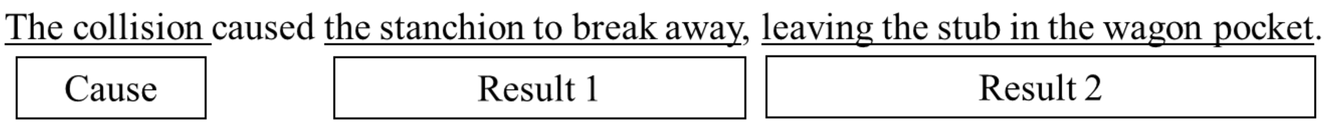


Figure 2-5: The concept of sentence-level causality extraction

Accident causality analysis studies in the context of the railway domain are still limited, and the result of the models designed in this context cannot fully provide insights into accidents. However, a small group of researchers in the context of chemistry have applied linguistic analysis in the model to analyse the causality of chemical accidents. Surprisingly, the result was improved significantly after several techniques of linguistic information extraction were considered (Du et al., 2020). In the context of information extraction, Semantic Role Labelling

(SRL) and Semantic Dependency Parsing (SDP) are the most powerful concepts to handle the causality within sentences. SRL and SDP are demonstrated in Figure 2-6 and Figure 2-7. SRL aims to assign labels to words or phrases in a sentence that indicate their semantic role in the sentence, such as that of an agent, goal or result, whereas SDP focuses on identifying semantic relationships between words in a sentence that form a graph. However, although these linguistic techniques have the potential to extract causality from accident reports in an efficient way, the connection between machine learning and linguistic analysis is still weak. Furthermore, the use of the output of such linguistic approaches is still unknown in terms of causality analysis, which might result in inappropriate understandings on the meaning of the original reports. The overview of the literature in the context of the railway industry through NLP applications is shown in Table 2-9.

	The	collision	caused	the	stanchion	to	break	away	,	leaving	the	stub	in	the	wagon	pocket	.
cause.01	A0			A1		C-A1											
break.01				A1						AM-ADV							
leave.01											A1		AM-LOC				

A0: causers or experiencers, A1: patient, C-A1: result, AM-ADV: adverbial of the patient, AM-LOC: location of the patient

Figure 2-6: The concept of sentence-level causality extraction with linguistic analysis tool

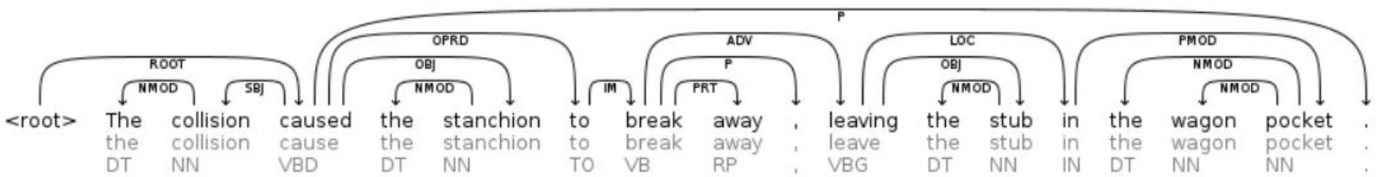


Figure 2-7: Demonstration of Semantic Dependency Parsing

Some criticisms relating to the use of NLP have emerged and been discussed in the literature. Among these criticisms, the major challenge remains the data quality and annotation. Many studies suffer from the restricted scale of labelled and consistent data for training the NLP model (Kim & Yoon, 2013; Kyriakidis et al., 2015; Zhou & Lei, 2018), resulting in limited insights and an inability to expand the scale of analysis. Another critical criticism is on the issue of data privacy and security (Worton, 2012). As railway safety data might contain sensitive information, legal and ethical obligations to protect the information from leaking to software providers must be prioritised to secure privacy. However, the abovementioned concerns could be addressed by leveraging unsupervised learning-based approaches to train the offline NLP model. Such an approach could avoid the heavy requirement of data annotation and restrict the data usage to individual devices rather than other servers. A more comprehensive method to secure data privacy is to follow the instructions designed by research ethics committees or praties, offering diverse perspectives of protecting data used.

Table 2-9: Overview of literature in the context of the railway industry through NLP applications

Purpose	Approach	Data		Result/ Performance	Source
		Input	Output		
Propose a taxonomy framework for railway accidents and a classification system	Text mining	Documents, n=392 (Written in Chinese)	Classification (6 clusters)	No performance of the model is revealed in the original paper	(Yu et al., 2018)
Classify the sentences in accident reports into two categories: accident description (AD) and causal analysis (CA)	Deep learning (M-CNN model)	Labelled sentences, n=1044(AD)/943(CA) (Written in Chinese)	Classification (AD or CA)	P*: 0.980, R*: 0.990, F1*: 0.985	(Hua et al., 2019)
Identify and extract the cause and result events (i.e., given the sentence is: A results in B, the algorithm needs to identify A (cause) and B (result) respectively)	Named Entities Recognition	Labelled sentences N=943(CA)	Classification (cause or result)	Cause: P*: 0.819, R*: 0.834, F1*: 0.826 Result: P*: 0.815, R*: 0.859, F1*: 0.812	
Propose a topic modelling approach and a classification model	Unsupervised Learning NER	Documents, n=298 (Written in English)	Classification (Factor of causes)	No performance of the model is revealed in the original paper	(Syeda et al., 2017)
Extract entities from accident reports	Named Entities Recognition	Documents, n=1066 (Written in Chinese)	Classification	Causes: P*: 0.7948, R*: 0.7339, F1*: 0.7664	(Li et al., 2018)
Cluster the cause of rail accidents	Supervised Learning	Documents, n=40164 (Written in English)	Classification (6 clusters)	F1: 0.71	(Heidarysafa et al., 2019)
Extracting entities from accident reports	Supervised Learning	Documents, n=120 (Written in Chinese)	Classification (12 clusters)	No performance of the model is revealed in original paper	(Wu et al., 2020)
Obtain evolution sequences of events in chemical accidents	Supervised Learning	Documents, n=5867 (Written in English)	Causality	F1: 0.85	(Du et al., 2020)

*P: Precision, R: Recall, F1: F1-score

2.3.6 Application of the ontology in the context of railway safety

Ontology is a branch of the knowledge management system, which aims to capture the domain knowledge of interests and support the interpretation of information by mapping the concepts logically. The railway safety domain has gained significant benefit from the application of ontologies. Due to the complexity of railway accident data, the literature widely uses the ontology approach to express the knowledge in the railway context for further analysis or decision-making (Cao et al., 2019; Debbech et al., 2020; Tutcher, 2015).

To obtain a valid ontology for specific objectives, the purpose of the ontology needs to be defined carefully, and the methods applied should be designed in a rigorous way. All ontologies stem from the upper-level ontology consisting of the general concepts that are commonly used in given domains, which are allowed to be used and represent other knowledge domains in similar structures (Debbech et al., 2020). The upper-level ontologies can also be considered as the starting point for formulating the definitions. On the other hand, the depiction-level ontologies developed from the upper-level ontology are built to demonstrate the specific knowledge or property contained in the upper-level ontology. The knowledge map can be linked to the actual data to provide the practical insights in the railway safety context through the development of the upper-level and depiction-level ontology.

Several ontologies are developed in the literature to represent the particular domain knowledge in the railway context, such as the lingual ontology for the purpose of extracting safety content from multi-lingual free-text safety incident reports (Hughes et al., 2019), the ontologies for supporting the capitalisation and exploitation of produced knowledge from accident records (Cao et al., 2019; Maalel et al., 2012b, 2012a), and for the purpose of a decision-making support system during the disruption (Wu et al., 2020). Other studies review the used ontologies to depict and aggregate the knowledge map (Hulin et al., 2016; Katsumi

& Fox, 2018; Tutcher, 2015). More studies applying the ontology in the railway context are shown in Table 2-10.

However, most railway safety data is recorded through unstructured text with solid description but with solid description on the sequence of events and the insights. The challenge of integrating data from multiple, unrelated sources into a unique framework for panel analysis results in the difficulties of exploring the knowledge in the context of railway safety (Katsumi & Fox, 2018). Before the development of the NLP technique, researchers needed to manually read the reports and extract the components to build the further model for horizontal analysis (Baysari et al., 2008; Wullems et al., 2013). The domain of interested knowledge is constrained in specific concepts, such as human error or infrastructure failure. To overcome the difficulty, some studies have applied NLP to handle the textual data. For instance, Figueres-Esteban et al. (2016) designed an NLP model to map the structure of reports and use the records Close Call System to determine the risky factors based on the frequency of the terms. Another example is Cao et al. (2020), who built an NLP model to transform unstructured records into the components in the pre-designed ontology and identify the risk source level of records.

Other issues of developing the ontology are the standardisation and the scalability. To enhance the reuse of ontologies, the Web Ontology Language (OWL) was established to specify the language and the formation used in building an ontology with the foundations in Description Logic (Baader et al., 2003). The axioms applied in Description Logic notation constrain the proposed taxonomy, which improves the readability and understandability of the designed ontology (Debbech et al., 2020). To discriminate the classes, properties and individuals in the ontology, the relationship is defined through notions of the subclass, conjunction, disjunction and negation, which allow us to categorise a particular object to a specific class based on the necessary and/or sufficient conditions (Katsumi & Fox, 2018).

Figure 2-8 shows an example of an OWL representation of vehicles. The sufficient conditions for the notions of the $HouseholdVehicle$ and the $Road–rail\ vehicle$ are illustrated. The OWL indicates that road-rail vehicles are vehicles which have access to part of the road and railroad systems, whereas household vehicles are not road-rail vehicles, and *vice versa*. Through such logic demonstration, the interested entities can be well recognised, and the inference becomes possible once new entities are introduced.

In terms of the scalability of the ontology, Grüninger et al. (1995) developed an ontology spectrum to illustrate the range of artefacts shown in Figure 2-9. The complexity of the expression on the specific domain increases toward the right end of the spectrum. The expressions at the left of the demarcation (syntactic, thesauri and metadata) only contain the concept of a particular individual without describing the relations. In contrast, the formal ontologies contain the element of inferring and reasoning, which allows the users to express the domain knowledge through a set of rigorous rules and help in automated reasoning of knowledge (also known as ontology learning) (Davies, 2010). However, the cost of complicated methods (the expressions on the right of the spectrum) is huge, and the scalability of the ontology would be restricted by either the size of the database or the performance of computational properties. Most previous studies fall in the Description Logic (OWL), RuleML (Rule Modelling Language) and SWRL (Semantic Web Rule Language) (Chimalakonda & Nori, 2020; Katsumi & Fox, 2018; Wu et al., 2020). Although the development of NLP has increased the momentum to move toward the right of the spectrum, the approaches are still controversial in the context of common logic (Asim et al., 2018; Shu et al., 2019), and the argument about recognising a computer's ability in understanding natural languages is continuing (Konys, 2018; Lehmann & Völker, 2014).

Table 2-10: Studies applying the ontology in the railway context

Purpose of using ontology	**Applied ontology**	**Data**	**Sources**
Extract safety content from multi-lingual free-text safety incident reports identifying specific classes of safety incident	Lingual ontology	5065 safety incident reports	(Hughes et al., 2019)
Using text network and graph database to map the structure of reports (like a word cloud with links)	None (only apply sentences segmentation)	150 records from Close Call System	(Figueres-Esteban et al., 2016)
Building risk ontology and scenario-risk-accident chain model (Integration of accident-risk ontology and context ontology)	Upper-level ontology + Depiction-level ontology	101 railway accident reports	(Cao et al., 2019)
Integrating the safety ontology for automobiles and railway vehicles from ISO 26262, EN 50126 and SIRF	Upper-level ontology	None	(Hulin et al., 2016)
Building a model to support the capitalisation and exploitation of produced knowledge retrieved from previous railway accident records	Depiction-level ontology	None	(Maalel et al., 2012a)
Mapping the potential parameters for building the ontology of accident scenario	Depiction-level ontology (event-oriented)	Used ontologies in the literature	(Maalel et al., 2012b)
Reviewing used ontology in the literature	Upper-level ontology + Depiction-level ontology	Used ontologies in the literature	(Katsumi & Fox, 2018)
Retrieving the information from reports and making a decision-supporting system	Upper-level ontology	120 metro accident reports	(Wu et al., 2020)

Table 2-10: Studies applying the ontology in the railway context (continued)

Purpose of using ontology	**Applied ontology**	**Data**	**Sources**
1 Transforming unstructured records into the knowledge base created by formal accident reports. 2 Identifying **risk source level** of records based on ranked formal reports through NLP techniques (classification).	Upper-level ontology + Depiction-level ontology	40,000 data records & 101 recorded accident cases	(Cao et al., 2020)
Proposing the instructions on integrating ontological data to monitor railway asset	Upper-level, domain, and application ontologies	None	(Tutcher, 2015)
Proposing a formal ontology to describe the impact made by a disruption to travellers' journeys	Depiction-level ontology	None	(Corsar et al., 2015)

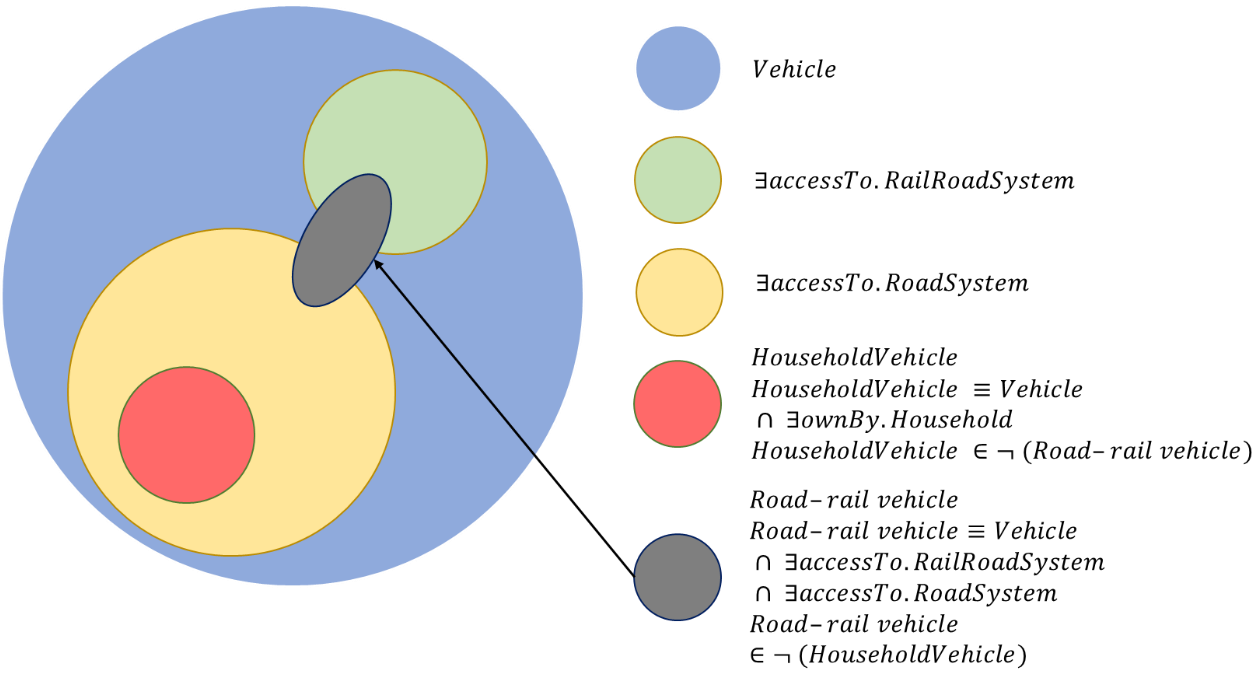


Figure 2-8: An example of an OWL representation of vehicles

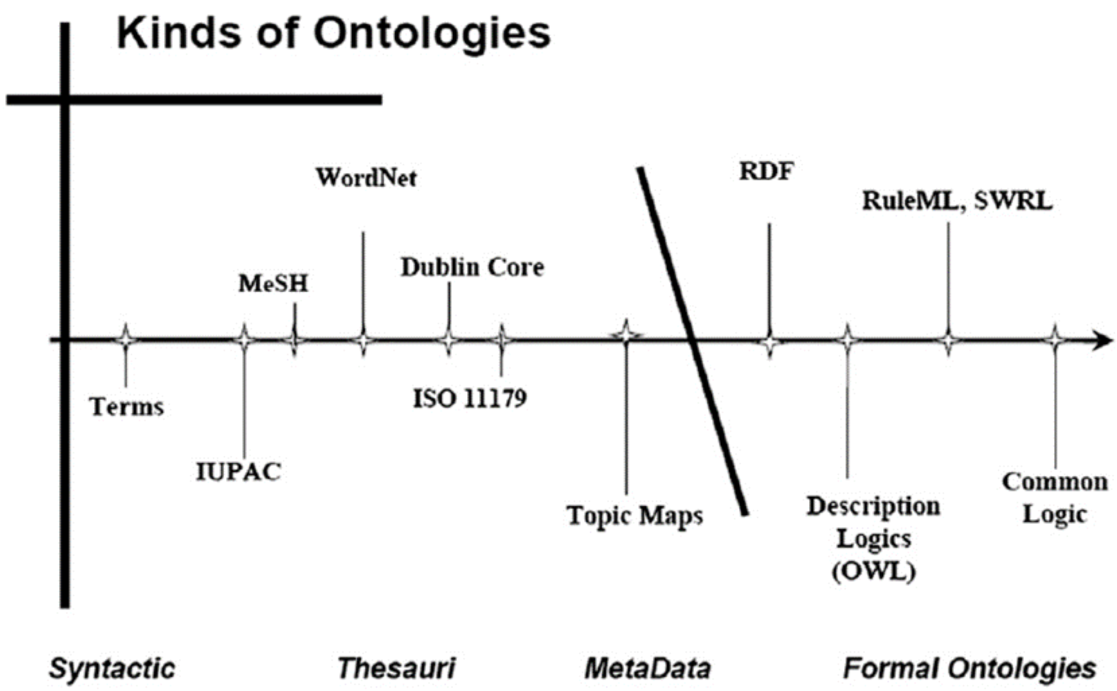


Figure 2-9: The ontology spectrum (Grüninger et al., 1995)

2.4 Synthesis of findings

This chapter reviewed how literature deals with the problem of assessing vulnerability in the railway industry. The concept of vulnerability was discussed, and a list of definitions was summarised in Table 2-1. It is concluded that the conflict of definitions in literature in the context of vulnerability assessment is due to divergent points of view. For instance, the definition of vulnerability varies for descriptive and operational purposes.

Subsequently, vulnerability assessment in the context of the transport system was reviewed in Section 2.1, and the foundational theory applied in this context was demonstrated. It is concluded that most of the studies have been completed on based on reliability theory and probability theory. Reliability theory mainly concentrates on reducing the vulnerability of a system by either increasing the resilience in preparedness or mitigating the impact after an accident. On the other hand, probability theory focuses on identifying the hazards and risky components in a system and revealing the possibility of the occurrence of each event. Hence, a considerable number of studies have been derived from either theory for different purposes. Several approaches have been designed to address the research questions under both theoretical bases, such as the statistical approach for research under probability theory, and the optimisation approach for research under reliability theory. However, through the increasing complexity of the socio-technical system, several theories have been built to understand the nature of the interaction, resulting in the unexpected outcome, within the system. Some valuable frameworks, like the SHELL and the HFACS models, have been built to interpret the incidents in a complex system, but the sequence of events and their probability remain unclear due to the difficulty of collecting analysable data. Moreover, according to the systems theory, interface weakness is one of the main factors that would lead the system to fail. However, the role that interface plays in the railway accident has not been fully understood in literature, which might make decision-makers underestimate the importance of interface and make the system vulnerable. Therefore, the first identified research gap from this review is the lack of understanding on causes and sequences of railway incidents.

On the other hand, the sequence of historical incidents has been well recorded by investigators in their jurisdiction. However, limited studies have applied such data to build a statistical model because of the difficulty of analysing textual data. Such limited knowledge on historical railway accident reports stems from a gap in the learning behaviour between railway accident investigators, which is the second gap identified in this thesis.

Additionally, a growing number of studies have put much emphasis on railway safety through state-of-the-art NLP techniques, which results in a significant opportunity to eliminate the restrictions on the analysis of big textual safety-related data. Some initial outcomes have indicated the possibility of letting an algorithm classify railway accidents based on the features of original accident records (Hadj-Mabrouk, 2020; Syeda et al., 2017, 2019). However, the extraction of critical hazards, which might trigger the actual accident, is still on the basis of human determination, which is time-consuming and labour intensive (Kim & Yoon, 2013; Zhou & Lei, 2018). Such limitations make the model unable to be updated after new railway accident cases or the development of technology. Nevertheless, the framework of required components is seldom set in previous studies (Li et al., 2018), which left the data unable to be reused and means further research has no standard to follow.

For this reason, another group of researchers has devoted much effort to building a standard for mapping the domain knowledge of their interests and representing that as an ontology. The ontology has been widely applied in other contexts, such as chemistry (de Matos et al., 2010; Krdzavac et al., 2019) and medicine (Arsene et al., 2011; Haendel et al., 2018). In the railway safety context, the ontology is conducted to express the lingual content (Hughes et al., 2019) and transform unstructured records into the components in the pre-designed structure (Cao et al., 2020). The ontology can provide a solid framework to indicate how the

NLP algorithm extracts the valid information from the original data and infers the relations between them. However, the combination of NLP and ontology applied in the railway safety context is rare, meaning the hazards not being identified efficiently. The ontology is not able to be updated based on the poor use of NLP techniques in the literature, making the result of ontology unable to be updated.

Furthermore, the recommendations in the accident investigation reports play an essential role in improving railway safety in practice. However, the issue of learning from recommendations across jurisdictions has never been researched before. Poor interaction between different railway safety agencies in different jurisdictions might result in severe railway accidents occurring, accidents which could have been prevented by following the recommendations made in other locations. To overcome the difficulties mentioned above, the research reported in this thesis designs several NLP models to process existing accident reports. Chapter 3 discusses the methodology developed to achieve this.

3. Research design and methodology

To automate the analysis of historical accidents and reveal the hazards in urban railway transport, Natural Language Processing (NLP) is used in this thesis to help analyse the textual information contained in the accident reports. This chapter provides the details of task identification and the methodology considerations, reviews and evaluation. The tasks required for answering the research questions are discussed and defined, followed by the introduction of potential candidate models of NLP, the introduction to the ontology design and the knowledge graph, and the evaluation of entity linking. The structure of the chapter is as follows: the research design process is designed (Section 3.1) and relevant tasks are defined (Section 3.2), followed by the introduction of Natural Language Processing (Section 3.3). Ontology design and the knowledge graph development are discussed (Section 3.4), and the synthesis of findings in this chapter is presented (Section 3.5).

3.1 Research design process

To address the research questions of this study, a mixed-methods design including both quantitative and qualitative approaches is proposed to provide a comprehensive understanding of railway safety across countries. Figure 3 1 depicts the research design process divided into a sequence of structured stages. The process begins with tasks definition (Section 3.2), ensuring the objectives of scope are closely aligned with the research questions. This is followed by data collection and pre-processing (Section 5.1 and 5.2), involving the sampling and preparation strategies. Subsequently, the following quantitative analysis involving NLP (Section 3.3) is separated into two parts: the BERTopic model and STM model (Section 4.1.1). Both approaches serve roles in analysing the main content and the recommendations of railway accident reports (See Figure 6-15). The quantitative analysis ends with the ontology and knowledge graph building (Section 3.4) which constructs a structured representation of the knowledge extracted from the data.

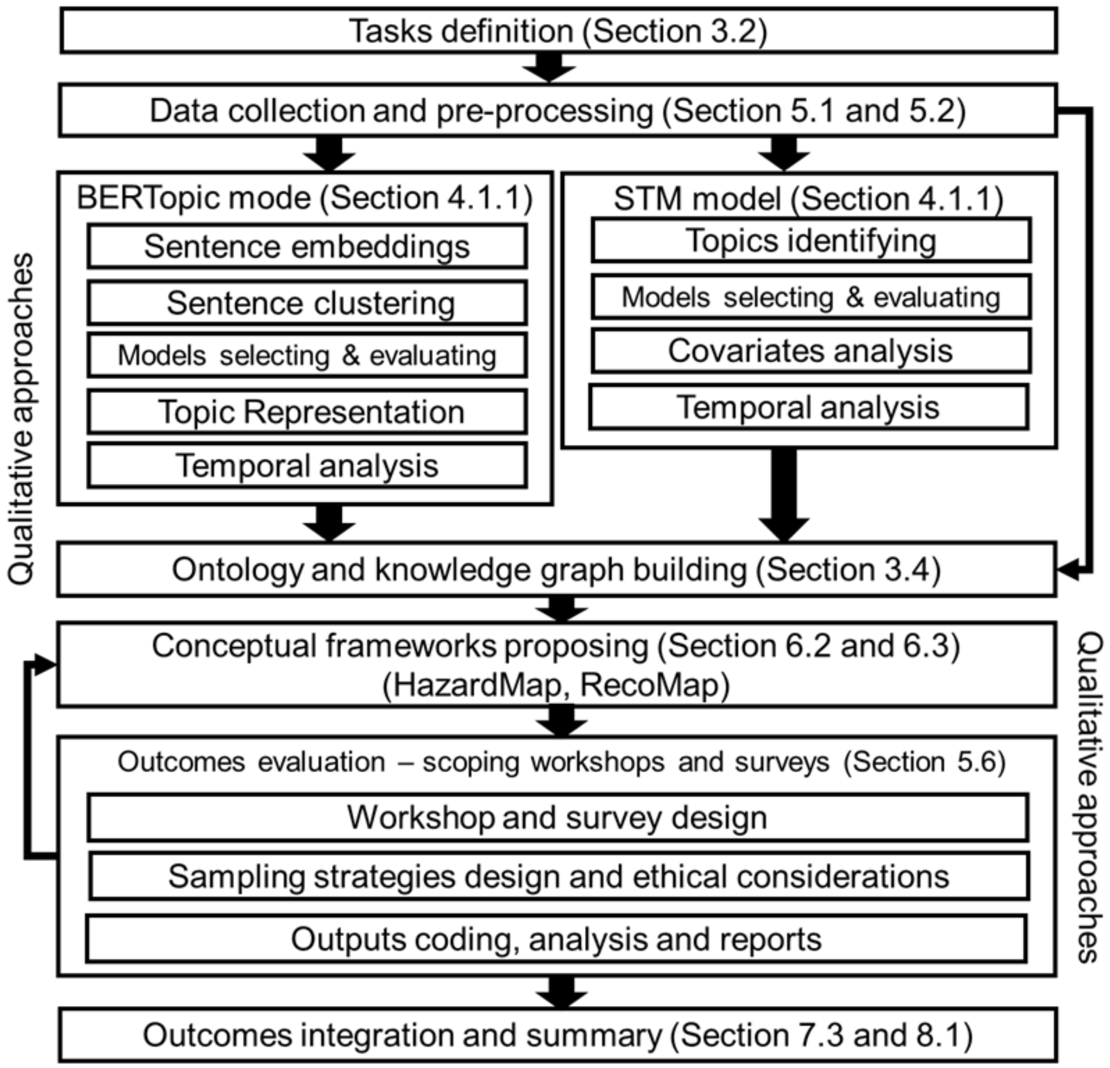


Figure 3-1: An overview of methodological approaches undertaken

Next, the outcomes of quantitative analysis support the development of conceptual frameworks (Chapter 6), building the schematic representations and interpreting complex relationships between concepts extracted from the railway accident reports. The conceptual frameworks and NLP models are evaluated through scoping workshops and surveys. These qualitative approaches aim to connect the outcomes to the practice by engaging with practitioners in the railway safety industry. The feedback and comments collected are used to further refine the conceptual frameworks proposed and ensure the research outputs are closely aligned with the practical railway safety operation. Finally, the outcomes integration and summary integrate the quantitative and qualitative findings and discuss potential underlying hazards within railway industry across jurisdictions (Chapter 7).

3.2 Definitions of tasks

To address research gaps discussed in the previous chapter, tasks and expected outcomes need to be identified before designing the methodological framework. First, an overview of the types of hazards is required. The model should be able to automatically identify critical hazards from railway accident reports (RQ1-1). The relations between hazards also need to be recognised for the purpose of understanding the mechanisms of hazards in each railway system, such as the impact of the failure of infrastructure or human behaviours on railway systems in different jurisdictions (RQ1-2). Next, the vulnerability that hazards cause across countries should be analysed and examined (RQ1-3). The different use of terminology should be addressed to ensure that all retrieved hazards link to correct entities in the developed knowledge (entity linking). For instance, the terms “level crossing” and “grade crossing” are expected to be linked to the same entities in the model for cross-country analysis (RQ1-3).

A classification system for the recommendations made by investigators is required to understand how solutions are proposed to address hazards from the socio-technical perspective (RQ2-1). The temporal factor should be included to evaluate the trend of the style of making recommendations and the behaviour of learning across time in the railway industry (RQ2-2). Last, opinions from practitioners are collected and analysed with the outcomes of topic modelling to understand whether railway accident report recommendations support the railway industry to learn across jurisdictions and time (RQ3-1). All results are integrated with the literature review to reveal the barriers to the railway industry learning across jurisdictions and time (RQ3-2).

In summary, the developed model needs to extract potential hazards from unstructured and unannotated railway accident reports from investigated countries. The model is also required to conduct sensitivity analysis, temporal analysis, covariate analysis and identification of accidents with similar hazards. An ontology is designed to store the retrieved hazards. A knowledge graph model is developed to address the issue of different terminology used across countries. Scoping workshops and a survey are conducted to understand the feedback about models developed and the opinions of learning from practitioners in the railway industry. An overview of required functions and outcomes for the research questions is shown in Table 3-1.

To satisfy the requirements, natural language processing, ontology development and the knowledge graph models are introduced in the following sections while the details of the scoping workshops and the survey are discussed in Chapter 4. Potential approaches using these techniques in the literature are reviewed, and a rigorous selection process is presented to determine methods used to build the model.

Table 3-1: Overview of required functions and outcomes toward research questions proposed

Research questions	**Analysis approaches**	**Expected outcomes**
RQ1-1: What is the difference in roles each hazard plays in various jurisdictions during railway accidents?	● Hazard identification ● Topic modelling (hazards) ● Sensitivity analysis	● Hazard lists ● Relative topics (hazards) ● Occurrence of hazards
RQ1-2: Do the same hazards occur in different jurisdictions and across time?	● Correlation analysis ● Temporal analysis (hazards)	● The trend of each hazard ● Correlation of hazards
RQ1-3: Do those hazards result in similar vulnerabilities in different jurisdictions and times?	● Covariate analysis ● Identification of similar accidents ● Ontology development ● Knowledge graph model	● Correlation between hazards and organisations ● Similar cases across time ● Railway accident ontology ● Entity linking model
RQ2-1: How do recommendations made by railway accident investigators address hazards identified from the socio-technical perspective?	● Topic modelling ● Systems theory	● Relative topics ● Distribution of each topic on the socio-technical framework
RQ2-2: Is there a transition in the style of making railway accident recommendations in each jurisdiction over time?	● Temporal analysis (recommendation)	● The trend of each recommendation topic
RQ3-1: Do railway accident report recommendations support the railway industry to learn across jurisdictions and time?	● Topic modelling (recommendation) ● Scoping workshops and survey	● Relative topics (recommendation) ● Opinions from practitioners
RQ3-2: What are potential barriers to the railway industry learning across jurisdictions and time?	● Interpretation of findings from this research supplemented by additional evidence	● A list of potential barriers and emerging underlying hazards

3.3 Introduction to natural language processing

Natural language processing (NLP) is a research field aiming to build the connection between human beings and computers in natural language. The main concern of NLP is to make the computer understand human languages and enable human–computer interaction (Gudivada & Arbabifard, 2018). The origins of NLP can be tracked back to the 1950s (Nadkarni et al., 2011), of which the purpose is to understand a large amount of textual data in an efficient way based on statistical techniques. Subsequently, the demand for text information retrieval increased and urged researchers to consider using artificial intelligence (AI) in analysing textual information. Hence, several NLP tasks were developed to response to industry, academia and government requirements. From the 1950s to 1970s, researchers have designed several approaches to solve some simple NLP tasks including Chomsky's 1956 theoretical analysis of language grammar (Chomsky, 1956) and Backus-Naur Form (BNF) notation (Aho et al., 1963) for identifying context-free grammar in articles. Another notable example is the concept of tokenisation which first raised in the 1970s (Nadkarni et al., 2011), aiming to propose a model that transforms text into tokens and generate a lookup dictionary. However, most methods before the 1970s required heavy human intervention and labour.

NLP experienced a great revolution in the 1980s after the introduction of machine learning (ML) methods (Zhang, 2014). The ML method is prominent for the ability to deal with probability questions, and the hand-written rules used before the 1980s associated the probabilities directly by building a series of constraints in the algorithms. Another significant breakthrough was the construction of large, labelled bodies of text (corpora), which was used to finetune the parameters in ML algorithms according to foundational statistical approaches (Nadkarni et al., 2011). The application of ML also raised several critical concepts like decision trees and feature-vector (Salzberg, 1994).

Researchers combined linguistics and AI with the advent of big data to enable computers to understand the inherent complexity and causality in natural language (Gudivada & Arbabifard, 2018; Krishnan & Rogers, 2015). To achieve this objective, several standardised sub-tasks in NLP were proposed for researchers to explore. Before the 2000s, the main NLP tasks were generally at text-level analysis, such as part-of-speech (POS) tagging, tokenisation, problem-specific segmentation, spelling and grammatical error identification and recovery and named entity recognition (NER) (Nadkarni et al., 2011). The ML techniques have significantly impacted most of these tasks by introducing shallow models such as support vector machine (SVM) and logistic regression. The shallow models use the concept of word embeddings and transform the text into the numerical matrix with high dimensional features (Young et al., 2018). The features contain linguistic information, such as grammar, phrases and slang. The performance of ML on NLP tasks is overwhelming but also comes with drawbacks, such as the requirement of hand-crafted features and time-consuming and extensive annotation, high error-susceptibility and limitations on understanding the sentence-level causality (Magnini et al., 2020; Mathews, 2019; Young et al., 2018).

In recent decades, the development of deep learning (DL) architectures has been significant through the rapid growth of computer hardware. The trend is catalysed by the success of complex conceptual methodology, including word embeddings, the recurrent neural network (RNN) and multi-level features engineering (Mikolov et al., 2010; Young et al., 2018). Because of the automatic feature representation learning, the DL framework gradually replaced most state-of-the-art ML approaches in several text-level NLP tasks (Collobert et al., 2011). Such advancement enables more difficult NLP tasks for researchers in either the computer science or linguistics context, including question-answering (Devlin et al., 2018), causality extraction (Fischbach et al., 2020; Khetan et al., 2020), language inference (Devlin et al., 2018) and language generation (Mathews, 2019). Hence, the paradigm in the NLP context shifted to build complex deep learning-based algorithms to solve complex NLP tasks. In recent years, a number of conceptual solid deep learning frameworks, such as Sequence to Sequence framework, attention mechanisms and long short-term memory (LSTM), have been proposed, which improve the performance of the NLP task-oriented language models and significantly reduce the computational complexity (Devlin et al., 2018; Magnini et al., 2020).

Both ML and DL treat the textual information as a numerical analysis by transforming the text into a high-dimensional numerical vector. The difference between ML and DL is the number of architectures in the model. The ML framework only contains one data processing layer without any neural network, whereas the DL framework includes several neural network architectures with multiple data processing layers, like a convolutional neural network (CNN) and RNN (Dang et al., 2020; Howard & Gugger, 2020).

On the other hand, the railway accident reports contain multiple dimensions, such as the description summary, causal chain analysis, accident analysis and recommendations. All contents are unstructured and unlabelled, and reports published in different jurisdictions may have other structures. The following sections review and compare off-the-shelf language models and training methods to select the appropriate approach that can address heterogeneous data and provide the required functions.

3.3.1 Building the language model – Let the computer read the text

This section reviews the language model for solving the NLP tasks. The evolution of how a machine understands natural language has advanced dramatically in recent decades, especially after the introduction of DL techniques. To ensure the proposed methodology in this thesis leverages state-of-the-art models, the language model's history is introduced first, followed by critical mechanisms and techniques.

The development of language models may be divided into the following periods: the application of statistical approaches (one-hot representation), word embeddings (distributional similarity-based representations), and language models (contextualised/ dynamic word embeddings).

3.3.1.1 Statistical approaches

In the early NLP development stage, the meaning of each word was delivered by using the bag of words (BoWs) or one-hot encoding (Harris, 1954). The fundamental method of the distributional structure mentioned is to assign a representation for each word by counting the occurrence of each word in a pre-set BoWs, containing a range of elements. For instance, assume a dataset *d* = ["I love NLP", "She loves NLP", "NLP requires several techniques"]. After the pre-processing, the input data would be *d'* = ["I love nlp", "she love nlp", "nlp require technique"], and the BoWs would be *D* = ["I", "she", "love", "nlp", "require", "technique"]. Finally, the one-hot representation of each word is as shown in Table 3-2.

Table 3-2: The demonstration of BoWs

	I	*she*	*love*	*nlp*	*require*	*technique*
d'[1]	1	0	1	1	0	0
d'[2]	0	1	1	1	0	0
d'[3]	0	0	0	1	1	1
one-hot representation	**1**	**1**	**2**	**3**	**1**	**1**

Subsequently, the one-hot representation was criticised due to the absence of semantic differences between words. Identical equal weights are considered in each word regardless of the semantic meaning. Hence, the concepts of term frequency (TF) (Luhn, 1957) and inverse document frequency (TF-IDF) (Jones, 1972) were introduced to enable the one-hot representation to consider the weight of each word across the collection of BoWs.

The TF approach can calculate how many times a term is used in one document, which equals the frequency of a term divided by the total number of existing terms. The TF of a term in a report can be calculated by Equation 3-1.

$$TF_{t,r} = \frac{F_{t,r}}{T} \qquad \text{Equation 3-1}$$

where $F_{t,r}$ is the frequency of a term t in the report r, whereas T is the total number of terms in the report r. The TF approach is feasible only while using one report. The TF-IDF approach should be conducted when multiple reports are used, allowing the consideration of the importance of a term in the whole database as shown in Equation 3-2.

$$W_{t,r} = TF_{t,r} * (log\left(\frac{D+1}{D_t+1}\right) + 1) \qquad \text{Equation 3-2}$$

where D is the total number of reports, and D_t is the number of times the term t appears in the whole database. However, the original result from the previous process is still limited in terms of description (Heidarysafa et al., 2019). For instance, the original result might not recognise the term “train” and “rolling stock” as the same entity. The term “rolling stock” might be divided into two meaningless terms “rolling” and “stock”. Additionally, the weight given by TF-IDF cannot capture the actual word meaning. The lexical meaning of a word should contain the related words as well. To consider the relationship between words, the co-occurrence vector was proposed to contribute to the one-hot representation and convert text into a sparse matrix (Lund & Burgess, 1996). Although there are several drawbacks in both TF-IDF and co-occurrence vector, both techniques still provided a pivotal statistical measure for researchers and were popularly used until the development of word embeddings (Hakim et al., 2014; Neto et al., 2000; Soleimani et al., 2019).

3.3.1.2 Word embeddings

To overcome the obstacles derived from statistical approaches, word embeddings (or distributional vectors) were introduced and captured the meaning of a word by reading its neighbouring words and giving the word a low-dimensional vector (Young et al., 2018). The similarity can be measured by calculating vectors through approaches like cosine similarity. To gain the word embedding of the data, the word2vec method designed by Mikolov et al. (2013) is widely applied in natural language analysis.

The word2vec model contains two sub-models: continuous bag-of-words (CBOW) and skip-gram model. The CBOW calculates the meaning of the target word through the surrounding context words across a window of specific size *k*, whereas the skip-gram model predicts the context words surrounding the target word. All words in the dataset are offered a series of weights based on their features. The conceptual framework of word2vec is shown in Figure 3-2. Given the original input text is “I like natural language processing”, and the window size *k* is 2, the left part is the mechanism of CBOW, and the target in this demonstration is “natural”. Since the window size is 2, then the context words “I”, “like”, “language” and “processing” would be the basis of the definition of the target word “natural”. The right part, which is the mechanism of the skip-gram model, predicts the words surrounding the word “natural”. ML or DL would train the parameters in the projection layer, and the result matrix represents the vector of each word in the input dataset. Figure 3-3 shows the demonstration of the result of the training. Given that the words in this model have n features, each parameter represents the weight the target word has in that feature. The $1 \times n$ matrix is the high-dimensional vector of the target word.

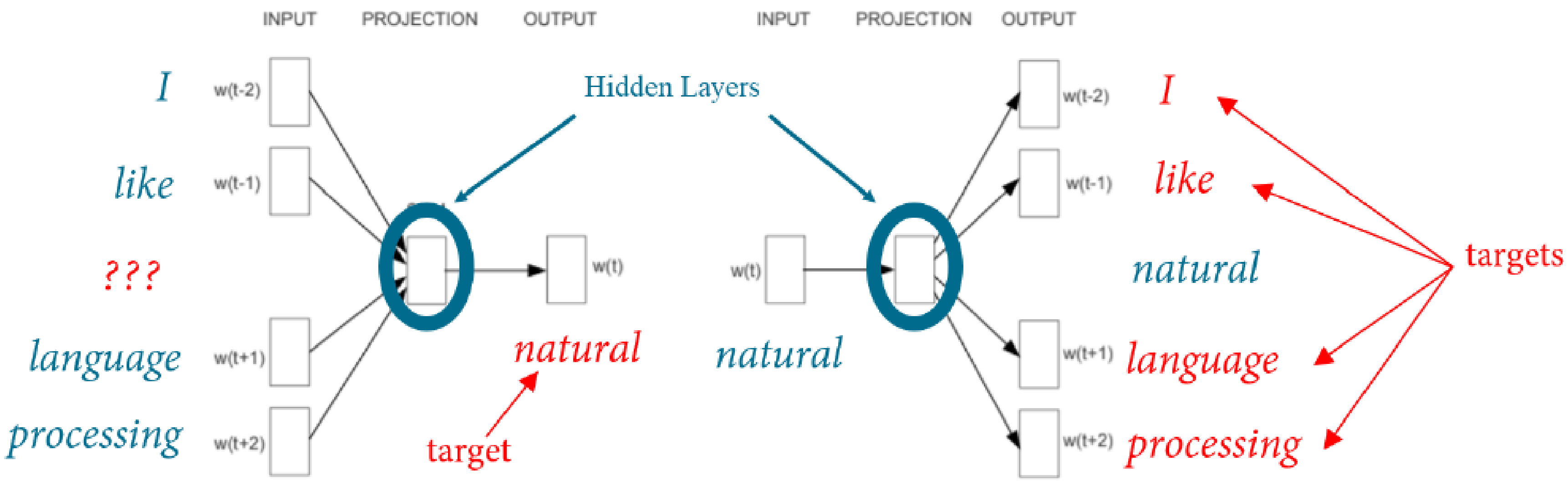


Figure 3-2: The conceptual framework of word2vec, source: Belkacem et al. (2017)

$$
\begin{array}{c} \\ \text{I} \\ \text{like} \\ \text{natural} \\ \text{language} \\ \text{processing} \end{array}
\begin{array}{c}
\begin{array}{ccc} f_1 & f_2 & f_n \end{array} \\
\begin{bmatrix} 0.1 & 0.05 & 0.02 \\ 0.8 & 0.01 & 0.02 \\ 0.4 & 0.2 \ \ldots & 0.05 \\ 0.1 & 0.01 & 0.02 \\ 0.02 & 0.02 & 0.03 \end{bmatrix}
\end{array}
$$

Figure 3-3: Demonstration of the projection layer (word embedding)

The conceptual framework of word2vec has inspired the design of the following language models. However, while this approach can identify the similarity of words, it cannot understand the relationship between the word and other sentences or documents, and the issue of co-reference resolution cannot be solved easily. The sequence of the sentence is not considered yet at this stage. Additionally, the sentiment of the word cannot be detected. For instance, the results of terms, such as positive and negative, sometimes have the same embedding (Young et al., 2018) due to the similarity of the neighbour words.

3.3.1.3 Language model (dynamic word embedding)

To overcome the issue of similarity raised by distributional vectors, the concept of recurrent neural networks (RNN) is introduced to NLP applications. The RNN is a type of neural network approach which bases the prediction on previous predictions (Elman, 1990). In other words, the result of word prediction would be “memorised” and becomes the basis of the prediction of the next word. Because the characteristic of memory allows capturing the inherent sequential nature of language, RNN is popularly used to deal with sequential NLP problems (Mikolov et al., 2010). As the demonstration shows in Figure 3-4, the RNN allows the model to learn the meaning of each word by sequentially reading the content.

The hidden layer of RNN contains accumulated crucial elements from previous steps. However, the parameters in the hidden layer would either explode or vanish while the width of RNN is increasing, also known as the vanishing gradient. Because each iteration in RNN refers to the error function with respect to the current weight, the weight would vanish after several multiplications and make the neural network unable to update the parameters in hidden layers. To handle this issue, two popular sub-networks are built into the RNN framework: long short-term memory networks (LSTM) (Gers, 1999) and gated recurrent unit (GRU) (Chung et al., 2014).

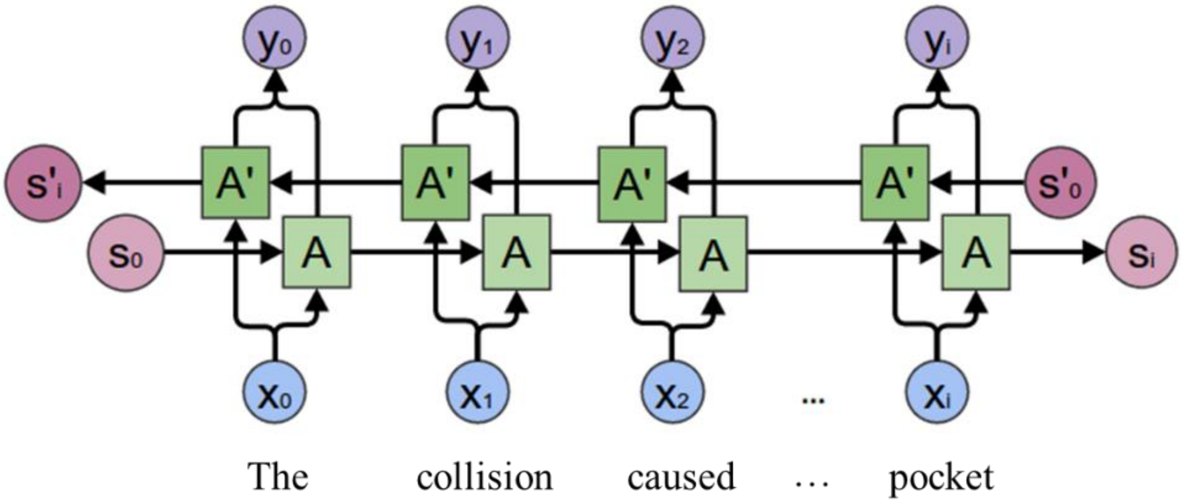


Figure 3-4: The conceptual framework of RNN (revised from Graves, 2013)

The LSTM (Figure 3-5) network contains three primary parameters: input gate, output gate and forget gate. The input gate determines whether the previous result would be input in this iteration; the forget gate determines whether the information should be programmed in this iteration, and the output gate determines whether the output should be passed to the next iteration or not (Hochreiter & Schmidhuber, 1997). Such a mechanism can avoid vanishing and exploding while processing the word embedding by controlling each gate's parameter.

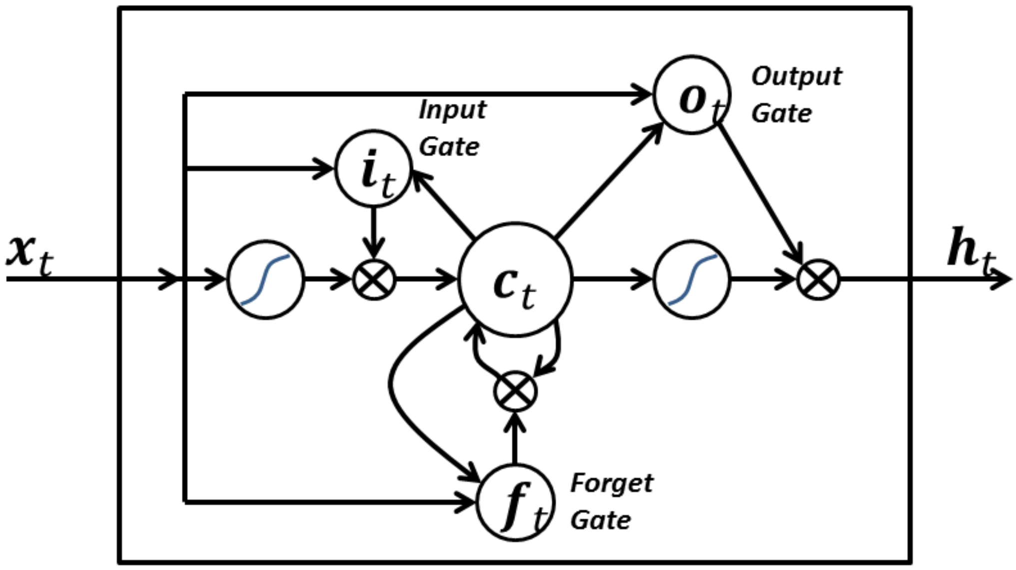


Figure 3-5: The mechanism of long short-term memory (LSTM) (Greff et al., 2017)

The GRU is the advanced version of LSTM, having most LSTM functions but only two parameters: update gate and reset gate (Cho et al., 2014). Figure 3-6 illustrates the mechanism of GRU where *r* represents reset and *z* represents update. The reset gate determines whether the input would be kept, and the update gate determines whether the output would be passed to the next iteration or not.

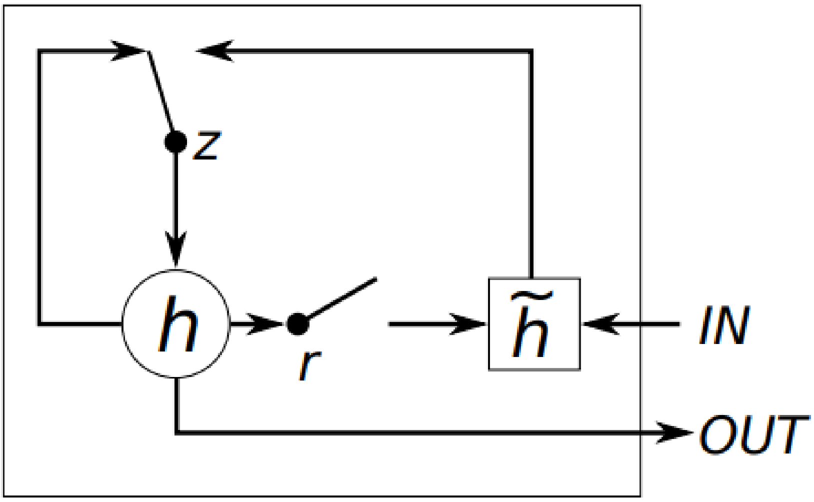


Figure 3-6: The mechanism of GRU (Chung et al., 2014)

The RNN applied LSTM and GRU cells in its network can improve the performance on understanding the sequential text considering the order of the words without the concerning vanishing gradient issue, which allows the model to understand the co-reference resolution and distinguish the different meanings of the same word in a different context. Figure 3-7 shows the mechanism of how the RNN model learns the word “strike”, which is similar to the word “collision” and “crash” in the same sentence.

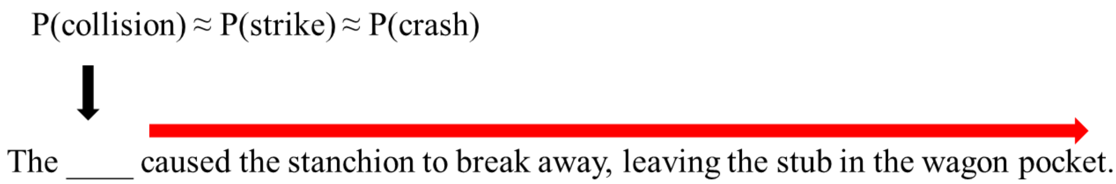


Figure 3-7: Illustration of how the RNN model applied LSTM or GRU analyses a word

Through significant improvement in computing power and computational techniques like parallelisation and deep learning, a new mechanism of programming "attention networks" was developed (Vaswani et al., 2017). Attention-based networks apply the linear combination of input and output tokens, allowing different lengths between the encoder and decoder (Bahdanau et al., 2015). Traditional word embedding-based models condense the original text in sentence-level analysis, creating significant bias if the input is too long or contains diverse topics (Young et al., 2018). The attention mechanism allows the output layer of the network to refer to the vectors of input data. The function of attention can be considered as a query and a series of key-value pairs to an output, and the queries (Q), keys (K), values (V) and the outputs are all single-dimensional vectors. The output is calculated by the weighted sum of the values, where the weights are based on the function of the query and the correlated keys (Vaswani et al., 2017).

The encoder-decoder architecture of the "Transformer" model proposed by Vaswani et al. (2017) uses Multi-Head Attention, which is a process performed by several Scaled Dot-Product Attention cells (*h* times, more details can be found in Vaswani et al. (2017)) in parallel with considering multiple dimensions of variables. The architecture of the Attention models is shown in Figure 3-8.

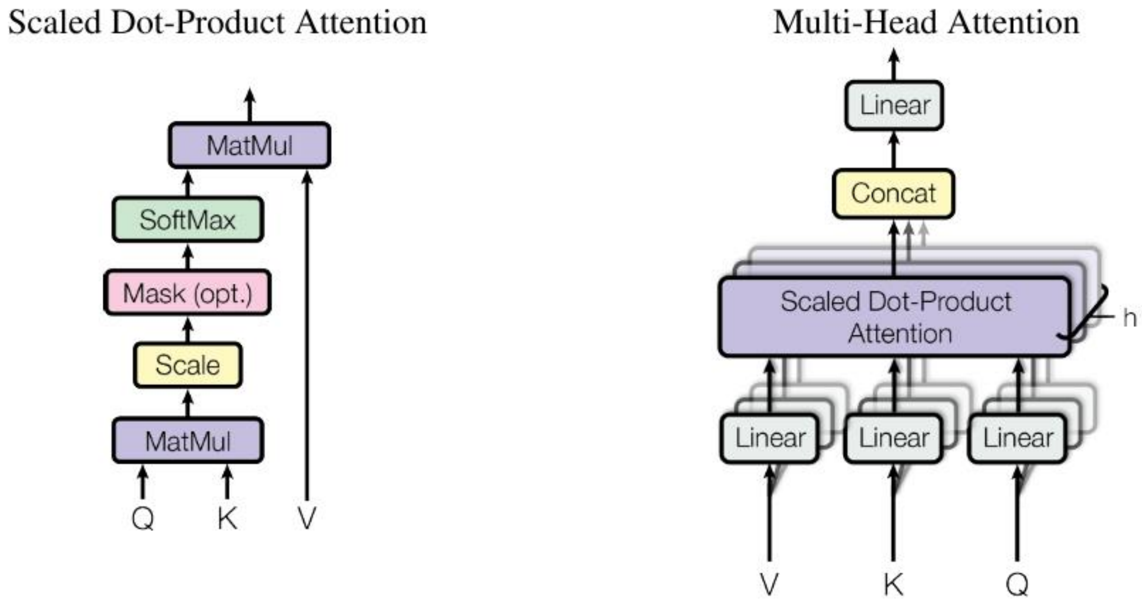


Figure 3-8: The mechanism of Scaled Dot-Product Attention (left) and Multi-Head Attention (right) (Vaswani et al., 2017)

The architecture of the Transformer is shown in Figure 3-9. The encoder of the Transformer contains multi-head attention and a full-connected feed-forward layer, and the decoder takes the output of the encoder in this iteration and an additional output of the decoder in the previous iteration as input to generate the final output of this iteration. The Transformer achieved state-of-the-art results in several NLP downstream tasks with other techniques and has been used in designing the framework of the language models.

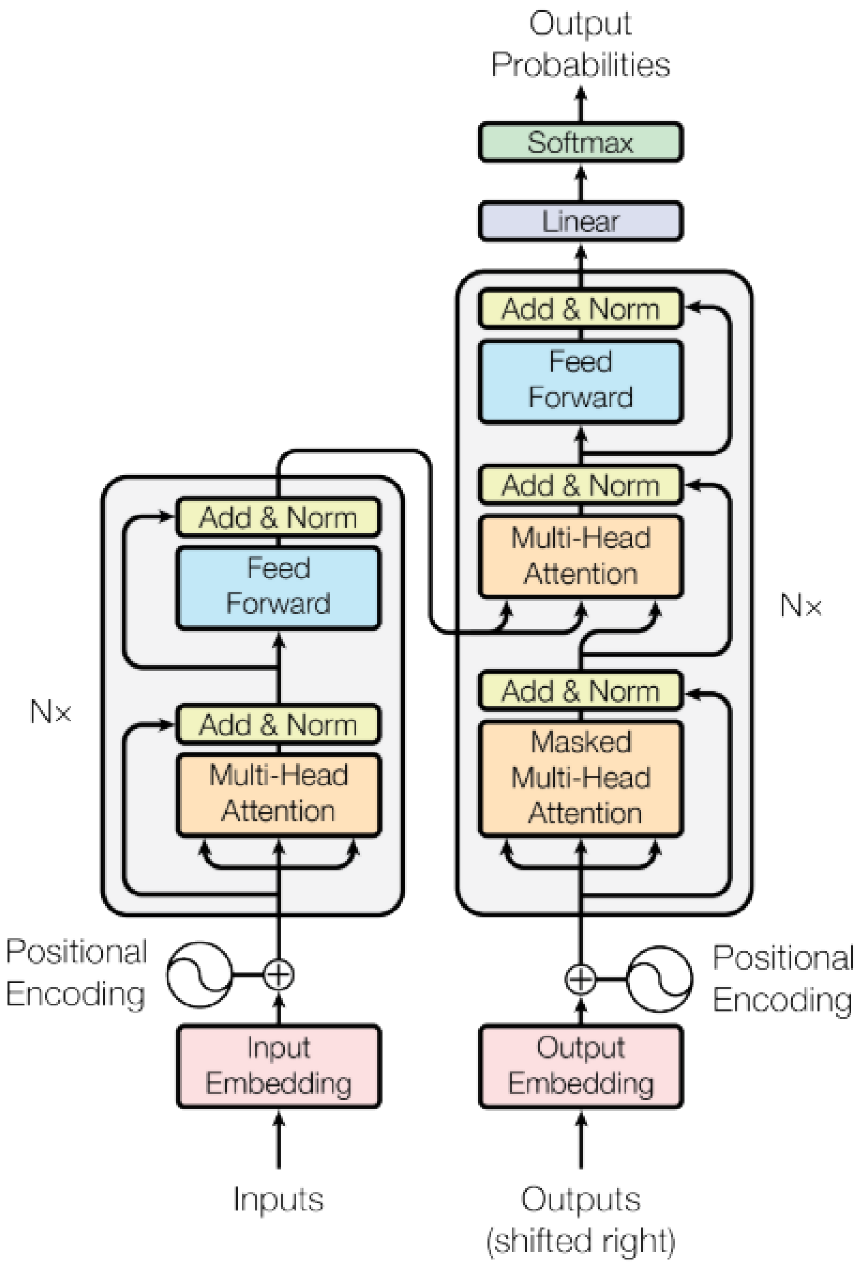
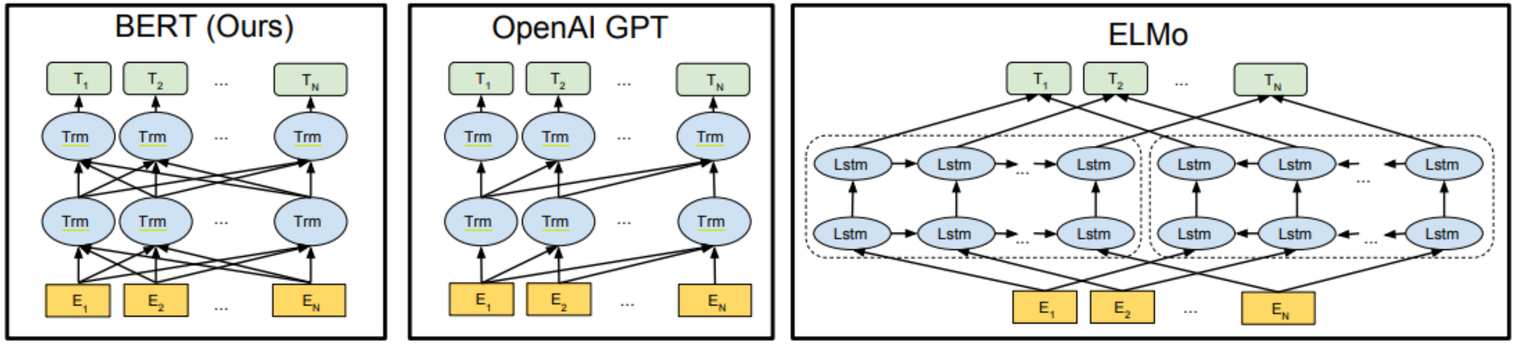


Figure 3-9: The structure of Transformer (Vaswani et al., 2017)

Since the development of deep learning, all mechanisms are allowed to work in the same architecture model, which allows the researcher to build a language model that can handle NLP downstream tasks with the best result. Several popular language models are designed using RNN architecture and LSTM, GRU and Transformer cells, such as embeddings from language models (ELMo) (Peters et al., 2017), OpenAI generative pre-trained transformer

(OpenAI GPT) (Alt et al., 2019) and bidirectional encoder representations from transformers (BERT) (Devlin et al., 2018). The deep network of each language model is shown in Figure 3-10.

The performance of each language model is examined with several corpora and benchmark databases. The BERT model has been considered the best model so far in dealing with the NLP problem (Devlin et al., 2018). BERT is a language model which Google developed, and consists of the techniques of deep learning, Transformer, machine learning and neural network. The BERT is a pre-trained model applying the masked language model (MLM) and next sentence prediction (NSP). MLM refers to the meaning of each word is given by other content in the document, and Figure 3-11 shows the mechanism of how BERT recognises a word. The masked words in both sentences are “strike”, and the BERT model calculates the most relevant word based on other texts in the sentence. The result demonstrates that the model can tell the difference between the word “strike” in a different sentence, implying that the polysemy issue (one word with multiple but related meanings) can be overcome through the masked language model. On the other hand, next sentence prediction means the probability that sentence will be connected to another sentence. The main task of next sentence prediction is to capture the relationship between two sentences. According to Devlin et al. (2018), the next sentence prediction task is specifically choosing two sentences and enabling the BERT model to predict whether another sentence actually follows one sentence.



*Trm: Transformers

Figure 3-10: Differences in pre-training model architectures, source: Devlin et al. (2018)

The resulting (**<u>strike)</u>** quickly grew and was joined by men from the nearby iron mills and factories.

```
Top 1 ( 5%): ['[CLS]', 'The', 'resulting', 'group', 'quickly', 'grew', 'and', 'was', 'joined', 'by'] ...
Top 2 ( 5%): ['[CLS]', 'The', 'resulting', 'company', 'quickly', 'grew', 'and', 'was', 'joined', 'by'] ...
Top 3 ( 4%): ['[CLS]', 'The', 'resulting', 'gang', 'quickly', 'grew', 'and', 'was', 'joined', 'by'] ...
```

An earthquake (**<u>strikes)</u>** the island and results in several derailment.

```
Top 1 ( 9%): ['[CLS]', 'An', 'earthquake', 'covers', 'the', 'island', 'and', 'results', 'in', 'several'] ...
Top 2 ( 3%): ['[CLS]', 'An', 'earthquake', 'on', 'the', 'island', 'and', 'results', 'in', 'several'] ...
Top 3 ( 3%): ['[CLS]', 'An', 'earthquake', 'reaches', 'the', 'island', 'and', 'results', 'in', 'several'] ...
```

Figure 3-11: Demonstration of masked language model (MLM)

The original BERT had two versions: the base version contains 110 million parameters with 12 Transformer blocks, 768 hidden layers and 12 self-attention heads; and the large version contains 340M parameters with 24 Transformer blocks, 1024 hidden layers and 16 self-attention heads. The cost of training each model is enormous. Fortunately, both pre-trained models are available online for developers to solve different tasks. Several studies using BERT to solve NLP problems with better performance have been published in recent years (Ganesh et al., 2020; Khetan et al., 2020).

The input embedding of BERT is shown in Figure 3-12. To ensure the BERT model can handle various downstream tasks, the input representation considers the input data's token embeddings, segment embeddings and position embeddings along with the identification tokens [CLS] and [SEP], which are used as the aggregation of sequential representations. The WordPiece embeddings with a 30,000 token vocabulary are applied in generating the token embeddings (Wu et al., 2016).

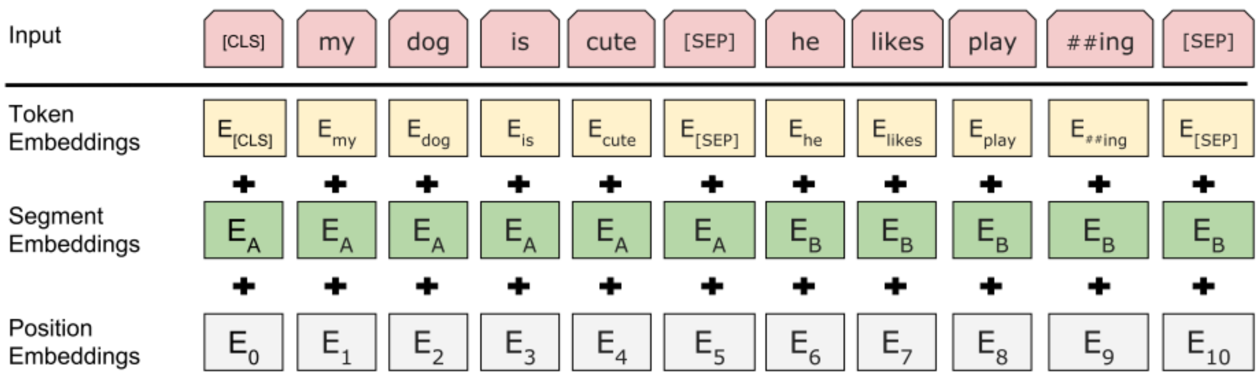


Figure 3-12: The input embedding of BERT (Devlin et al., 2018)

Since the powerful open-resourced model BERT was released in 2019, researchers have started to use it to overcome previous limitations in solving NLP problems. For instance, the concept of the MLM approach that BERT applies is extended to eliminate the requirement of labour-intensive data labelling. The MLM is able to find the top predicted words of the selected topic, and the predicted words become the basis of classification (Meng et al., 2020). The BERT model is also applied to handle the causality extraction task with considerable performance. The model designed by Khetan et al. (2020) on the basis of pre-trained BERT can detect the "Cause-Effect" sentence, the position of events in the sentence and predict the events. The corpus SemEval 2007 (Beamer et al., 2007), SemEval 2010 (Hendrickx et al., 2010) and ADE datasets (Gurulingappa et al., 2012) are used, all of which contain "Cause-Effect" annotations.

The subsequent sections illustrate the process of training an NLP model with different architectures and purposes.

3.3.2 Training an NLP model

To retrieve the knowledge of interest, a model must be trained to convert textual data into machine-readable data. Several machine learning-based approaches are established in the literature: rule-based approach, supervised learning, semi-supervised learning, and unsupervised learning.

3.3.2.1 Rule-based approach

The rule-based approach aims to explore the features of natural language and set a number of linguistic patterns manually to extract explicit or implicit clues. The patterns can be generated by lexico-semantic analysis, syntactic analysis or customised design, depending on the interests of the research. This implies that the required components for training a model via the rule-based approach are patterns and rules rather than labelled data. For example, the snowball search algorithm is one of the classic rule-based approaches, which aims to extract relations by defining seed causative verbs and identifying all connected subjects and objects, then searching the corpus again to find the new causative verbs connecting to these subjects and objects again. All selected candidates are evaluated by similarity measurement and relation-pair with a similarity score over a specified threshold considered a true relation (Alashri et al., 2018; An et al., 2019; Heindorf et al., 2020). By applying such approaches and the assumption, the causation can be concluded as a sequential flow to elaborate on the nature of a specific type of accident.

Although building up comprehensive patterns and rules is extremely time-consuming, in a specific context, the rule-based approach usually leads to higher performance than the machine learning approach. Given that the rule-based approach does not require labelled data and railway accident reports or records are highly homogeneous in terms of terminology and writing style, such an approach has strong potential in the railway accident causation analysis context.

3.3.2.2 Supervised learning

Supervised learning uses labelled data (such as railway accident data with the types of causes as a tag on each datapoint) during the entire training process, and the output model will predict the unseen data based on the way that training data is labelled. For instance, a model to predict the expected number of railway disruptions is trained by feeding the model the historical railway disruption data and assigning the locations and duration of the railway disruption as predicted targets (Yap & Cats, 2019). In the railway accidents analysis context, the labels used can be whether a hazard is involved, or what correlations of hazards exist in an accident. However, several studies argue that training a supervised learning model can match or exceed human performance only when the training dataset has more than 10 million samples (Goodfellow et al., 2016; Yang et al., 2021), and the size of annotated data for causal relation extraction benchmark is still not satisfactory to train a powerful learning model. Complex causation tasks will cause a higher risk of failing to retrieve an acceptable model via supervised learning without adequate labelled data.

3.3.2.3 Semi-supervised learning

On the other hand, semi-supervised learning is trained with only a small amount of labelled data and a large amount of unlabelled data. Because only limited labelled data is available, several assumptions or pre-defined relations borrowed from rule-based studies are required to guide the model on prediction. For instance, Lee et al. (2021) divided the data into normal and abnormal conditions before training a model for predicting the structural integrity of a railway bridge. Despite taking advantage of the benefits of rule-based and supervised learning approaches, the design of the model becomes complicated and has seldom been considered in causation extraction tasks due to the heterogeneity of required data, including labelled data and rules.

3.3.2.4 Unsupervised learning

Unsupervised learning is trained on unlabelled data only, and the model is forced to classify the data by modelling topics and building imaginative content. The user needs to interpret the outcome based on the nature and background knowledge of input data. For instance, Lasisi and Attoh-Okine (2020) leveraged unsupervised learning and experts' knowledge to predict rail track geometry defects. The performance of machine learning models mostly depends on training data and the quality of labelled data. Despite no requirement of labelled data for training a model via unsupervised learning, the performance might not be satisfactory, and manual interpretation is needed (Xie et al., 2019; Yamashiro & Nonaka, 2019). Hence, researchers using this approach must have the domain knowledge to reason the results of unsupervised learning. Nevertheless, the causations might not be available by only training one model because only similar topics are extracted without further information, suggesting the relations between extracted topics need to be associated with each other manually.

3.3.3 The trade-off between approaches to railway accident analysis

To select the most feasible approach for conducting a railway accident analysis, the goals of research must be set, and the nature of railway accidents and the way reports are written must be understood. First, railway accidents have features of low occurrences but rich information in each datapoint, suggesting that sentence-level causation analysis should be preferred rather than document-level causation analysis unless either specific goals are set, or the number of data items is satisfactory (Fischbach et al., 2020). Second, each record might have individual chapters or sections with various purposes, which could be analysed separately. However, this issue can be addressed by using unsupervised learning approaches because the purpose of sentences in each chapter is expected to be significantly different. For instance, the purpose of the summary chapter of a report is to present the overview of the accident, whereas the recommendation chapter elaborates on the improvements for the railway industry.

Table 3-3 summarises the critical features of each approach. Different tasks and types of datasets significantly influence the selected approach. In the context of railway accident analysis, most datasets are recorded in an unstructured way. Therefore, the consistency of the data needs to be estimated before conducting a rule-based approach. The resources for annotation are another critical factor in determining if the (semi-) supervised learning based approach is applicable or not. An inappropriate decision might result in poor performance or unreasonable outcomes (Yang et al., 2021).

To sum up, supervised learning and semi-supervised learning approaches have strong performance on predicting unseen data and allow users to narrow the results down to a set of categories of interests. However, intensive labour is required to annotate the data and the performance of the model heavily replies on the quality of annotation and the quantity of data. The purpose of the model in this thesis focuses on extracting hazards rather than prediction annotated data is also not available, and the quantity of railway accident reports is not satisfactory. Therefore, supervised learning and semi-supervised training are not considered to be applied. On the other hand, a rule-based approach offers higher performance on consistent data and ability of reasoning and interpretation. The challenge lies in the obstacles of setting comprehensive rules for extracting hazards of interests and the constraints on the structure of input data. Because railway accident reports from four countries are used, significant type II errors might occur due to various lexicons and writing styles. Last, despite the risks of black box issues while training and the manual interpretation of outcomes, the unsupervised learning based approach addresses the difficulty of annotation and allows users to extract potential hazards from heterogeneous data without heavy human intervention. Thus, the unsupervised learning based approach is selected to train the NLP model. More discussion about the implementation of unsupervised learning and the risk management of interpretation and evaluation is presented in the next chapter.

Table 3-3: Overview of critical features for rule-based, (semi-) supervised learning based and unsupervised learning based approaches

	Rule-based approach	**(Semi-) Supervised learning based approach**	**Unsupervised learning based approach**
Method	Manually set a series of rules, such as lexico-semantic patterns, to extract the hazards from unstructured data.	Manually annotate data and let the machine determine patterns to identify hazards based on labels from unstructured data. Pre-setting rules for the computer to predict is possible.	Directly input data and let the machine determine patterns without seeing any labels to extract hazards. The combination of several models is possible to achieve the goal.
Data	● Pre-designed rules	● Structured/labelled data	● Unstructured/plain text
Pros	● The ability of reasoning and interpretation ● Being able to build a labelled corpus ● Higher performance on data in the same context ● Efficient and consistent ● Requiring less computational power and time	● The ability to identify unforeseen patterns ● The ability to predict the unseen text (in the same context) ● Less human intervention after finalising the model ● Flexibility to analyse new data ● Several publicly available off-the-shelf tools	● No requirements for data annotation ● The ability to analyse the unstructured text without labels ● Less human intervention while training ● Allowance of heterogeneous data ● Several publicly available off-the-shelf tools
Cons	● Requiring experts or professional knowledge to set rules ● Possibly missing or underestimating (type II error) ● Less flexibility to analyse new data	● Requiring (expert-) annotated training data ● Black box issues while training ● Higher requirements of data for complex hazards	● Heavy reliance on manual interpretation ● Black box issues while training

Table 3-3: Overview of critical features for rule-based, (semi-) supervised learning based and unsupervised learning based approaches (continued)

	Rule-based approach	**(Semi-) Supervised learning based approach**	**Unsupervised learning based approach**
Builds	● Manually design	● Manually design ● Off-the-shelf tools ● Knowledge databases	● Manually design ● Off-the-shelf tools ● Knowledge databases ● Combinations of models
Ref.	Alashri et al. (2018); An et al. (2019); Heindorf et al. (2020); Yang et al. (2021)	Goodfellow et al. (2016); Yang et al. (2021); Lee et al. (2021); Yap & Cats (2019)	Lasisi & Attoh-Okine (2020); Xie et al. (2019); Yamashiro & Nonaka (2019)

3.4 Introduction to the ontology design and the knowledge graph

The term “ontology” is borrowed from the philosophy that refers to the science that describes entities and their relations in the real world (McGuinness & Harmelen, 2004). In other contexts, ontology has been developed into a tool presenting the domain knowledge of their interests. Considering that the nature of railway accidents has been recognised as a linear-interacting and tight-coupling system (Shrivastava et al., 2009), the contributing factors in a railway accident involve several animate and inanimate components which are governed by multiple stakeholders and organisations.

In addition to the direct causes, the underlying factors like poor management policies and decisions are also held accountable in an accident for creating an unsafe operational environment as a catalyst. Accidents occur through a series of structured failures, negligence, or underlying environmental factors. The complicated and multiple causes of a railway accident make it difficult to understand the hazards and the triggers of the accident in a comprehensive way. Additionally, the ontology with explicitly defined semantics could help researchers to transcribe the unstructured text into machine-readable languages, which can support knowledge management and reasoning services on integration, data validation and inference (Katsumi & Fox, 2018). Hence, the ontology is imported to illustrate the map of railway accidents.

To describe the entities and the way they relate to each other, the ontology contains six core components: individuals, classes, attributes, relations, function terms, and axioms. The definition of each component is shown in Table 3-4. Consequently, a structured model is built to express the domain knowledge of the ontology consisting of the aforementioned critical components (Debbech et al., 2020). Let *O* be the ontology considered as a 6-tuple:

$$O = \{I, C, Att, R, F, A\} \qquad \text{Equation 3-3}$$

where $I, C, Att, R, F\ and\ A$ represent individuals, classes, attributes, relations, function terms and axioms, respectively. The purpose of the ontology is to be able to illustrate the nature of railway accidents in the real world and be extendable.

Table 3-4: Critical components and examples in an ontology (revised from Reyes-Peña & Tovar-Vidal, 2019)

Name	**Description**
Individuals	The base unit of an ontology, representing any concrete object or abstract individual. i.e., staff, train, passenger.
Classes (concept)	Representing a group of different individuals sharing common characteristics. i.e., train is the class of all trains, or any abstract which can be described by the criterion for being a train.
Attributes	Representing something expressing a fact that is specific to an object. i.e., train has a “train number”, a “conductor”.
Relations	Referring to the relationship between components in an ontology structure.
Function terms	Objects with the purpose of retrieving information from other objects.
Axioms	Consisting of restrictions, rules, or other logic correspondences definitions to ensure the ontology has valid structure and relationships.

3.4.1 The design of ontology

To design an ontology meeting the interests of stakeholders (i.e., the railway industry and the railway safety agencies), rigorous processes should be built to ensure a convincing result. Table 3-5 shows the process of modelling a domain ontology, containing phrases of specification, conceptualisation and evaluation.

Table 3-5: The process of modelling an ontology for a specific domain context (revised based on Brusa et al., 2006; Reyes-Peña & Tovar-Vidal, 2019)

Phrase	Task	Description
Specification	The ontology goal and scope	Defining the goal of the ontology and specifying the boundary
	Domain description	Describing the domain of the ontology
	Motivating scenarios and competence questions	Creating the scenario description for modelling the informal logic knowledge
	Ontology granularity	Deciding the number of attributes in the ontology
Conceptualisation	Extending the ontology by increasing data	Applying the designed algorithm to identify and classify the entities
	Re-identification of classes, relations and attributes	Reviewing and modifying connections between concepts
Evaluation	Verification	Determining whether the ontology meets the requirement of the competency questions defined in the specification phrase
	Validation	Determining whether the ontology can represent the knowledge of interests comprehensively

3.4.1.1 Specification

For the implementation of the specification, first the goal and the scope of the designed ontology should be clarified, followed by the determination of applied data and type of ontology. In the literature, experience-based and evolutive approaches are widely used to develop an ontology (Brusa et al., 2006). The experience-based approach aims to build conceptualised ontology via workshops or consulting subject matter experts (Tutcher, 2015) or based on proposed ontologies in the literature (Corsar et al., 2015; Katsumi & Fox, 2018; Maalel et al., 2012b). On the other hand, evolutive approaches develop ontologies with iterative processes and extend them with real-world cases. For instance, Cao et al. (2019)

developed an ontology for railway risk analysis and a scenario–risk–accident chain model by integrating present ontologies and extending them with railway accident cases; and Wu et al. (2020) developed a decision-making support system based on the case-based metro accident ontology.

In addition, developed ontology can be classified into four sub-modules depending on different purposes: upper-level ontology, domain-level ontology, depiction-level ontology and application ontology (Corcho & Gómez-Pérez, 2000; Sintek & Decker, 2002). An upper-level ontology aims to cover general entities and provide definitions of concepts at the abstract level. The upper-level ontology is the root of other ontologies and promises to be extended to domain ontologies by offering basic elements and their relations in the real world. On the other hand, domain-level ontology further illustrates how entities in a field of interest relate to each other. For instance, a railway accident domain ontology should be able to elaborate on the causality between trains, risks, hazards and other entities (Cao et al., 2019). To extract such causality, depiction-level ontology is required to bridge the gap between domain-ontology and data. The depiction-level ontology classifies entities extracted from data into a map and describes factors involved in the domain of interest. For instance, the railway organisations participating in a railway accident are expected to be identified and associated with the built ontology. Last, the application ontology is the final product of the developed ontology and is prepared to be used in other projects or research. A review of application ontologies is recommended to understand the existing knowledge framework and determine whether to reuse and extend these ontologies (Hulin et al., 2016).

A review of existing developments of ontologies in the railway context is synthesised (Table 3-6). Most upper-level ontologies[1] are based on subject matter experts' knowledge, and only

[1] Upper-level ontology: a general ontology that describes the sequences of an accident in a temporal way.

limited depiction-level ontologies[2] are proposed due to highly labour-intensive analysis. Despite some trials to implement NLP to automate the process of ontology generation, the issue of heterogeneous data languages remains unaddressed in the development of general ontology (Cao et al., 2020; Hughes et al., 2019). Nevertheless, merging the published ontologies in Table 3-6 for integrating railway risk knowledge is extremely challenging due to the absence of uniform rules and protocols. These obstacles might hinder researchers and practitioners from understanding and extending state-of-the-art knowledge.

[2] Depiction-level ontology: an ontology that describes things involved in an accident (sometimes with solutions or recommendations).

Table 3-6: Overview of the development of ontology in the railway context

Purpose of using ontology	Applied ontology	Data	Sources
Extract safety content from multi-lingual free-text safety incident reports to identify specific classes of safety incident	Lingual ontology	5065 safety incident reports	(Hughes et al., 2019)
Using text network and graph database to map the structure of reports	None (only apply sentences segmentation)	150 records from the Close Call System	(Figueres-Esteban et al., 2016)
Building risk ontology and scenario-risk-accident chain model (integration of accident-risk ontology and context ontology)	Upper-level ontology and depiction-level ontology	101 railway accident reports	(Cao et al., 2019)
Integrating the safety ontology for automobiles and railway vehicles from ISO 26262, EN 50126 and SIRF	Upper-level ontology	Subject matter experts	(Hulin et al., 2016)
Mapping the potential parameters for building the ontology of accident scenario	Depiction-level ontology (event-oriented)	Used ontologies in the literature	(Maalel et al., 2012b)
Reviewing used ontology in the literature	Upper-level ontology and depiction-level ontology	Used ontologies in the literature	(Katsumi & Fox, 2018)
Retrieving the information from reports and making a decision-making supporting system	Upper-level ontology	120 metro accident reports	(Wu et al., 2020)
Transforming unstructured records into the structured knowledge created by formal accident reports and identifying risk source level of records based on ranked formal reports through NLP techniques (classification)	Upper-level ontology and depiction-level ontology	40,000 data records & 101 recorded accident cases	(Cao et al., 2020)
Proposing the instructions on integrating ontological data to monitor railway assets	Upper-level, domain, and application ontologies	Subject matter experts	(Tutcher, 2015)
Proposing a formal ontology to describe the impact made by a disruption on travellers' journeys	Depiction-level ontology	Used ontologies in the literature	(Corsar et al., 2015)

3.4.1.2 Conceptualisation

To ensure developed ontology covers comprehensive knowledge in scope, the re-identification of classes, relations and attributes is required to improve the structure of the ontology by integrating concepts and refining existing ontologies (Kotis & Vouros, 2005). First, an upper-level ontology is introduced and further finetuned to a domain ontology by consulting and reviewing domain knowledge. Subsequently, the entities in the railway accident reports should be linked to corresponding concepts as instances. For instance, the entity “track” is expected to be connected to the concept “rail infrastructure”. A model for the iterative process of entity identification and linking is designed and illustrated in Sections 3.4.2 and 3.4.3. Next, connections at each level are reviewed and modified by merging or extending concepts and their instances. An example of relationships between levels in an ontology is shown in Figure 3-13.

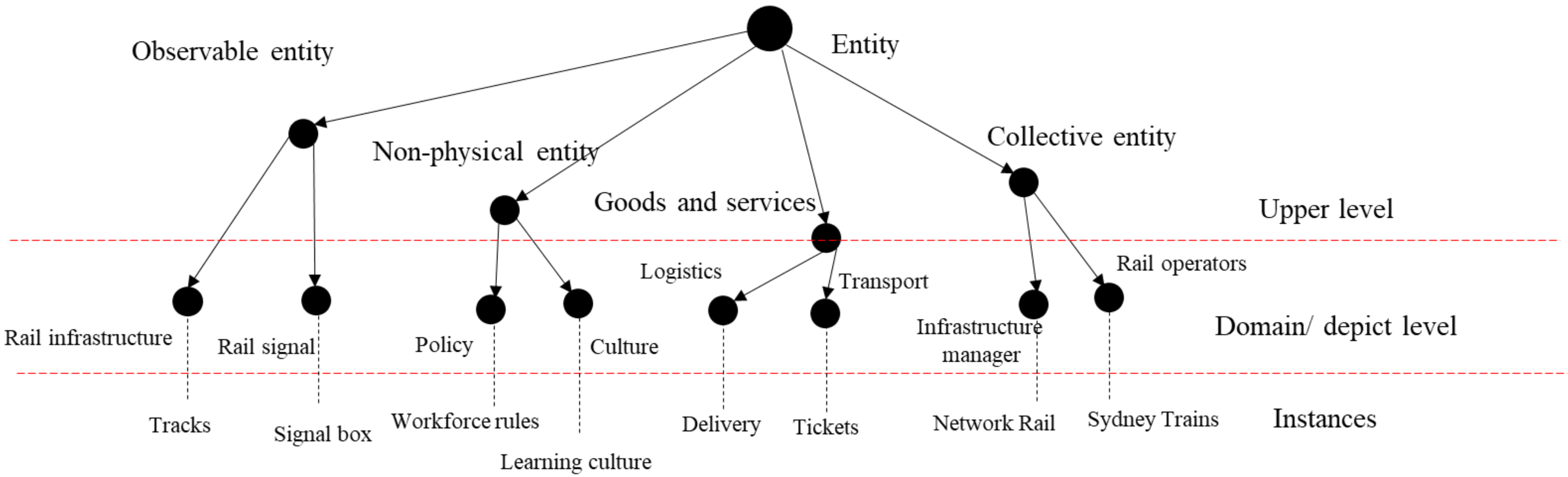


Figure 3-13: An example of ontologies, concepts and instances

3.4.1.3 Evaluation of the ontology

The evaluation consists of two parts: verification and validation. The verification aims to ensure the definitions used in the ontology are implemented correctly in the real world and three factors are examined manually: inconsistency, incompleteness and redundancy (Lovrencic & Cubrilo, 2008). Inconsistency focuses on whether the structure of the ontology and the definitions of each concept meet the logical requirements. Incompleteness refers to the completion of entities, concepts and their relations for representing the knowledge of interests. Redundancy examines the ability of ontology to discriminate entities with multiple definitions or different types of terminology to the same concept. On the other hand, validation ensures that the created model is compliant with the real world. Lovrencic and Cubrilo (2008) suggest that an application ontology concentrating on real-world cases should be developed and that comparing the created ontology and application ontology indicates whether the real-world cases are fully illustrated.

3.4.2 Knowledge graphs

To address the heterogeneity of railway terminology used in different countries, knowledge graphs are introduced to standardise extracted mentions and convert them to the correct entity before connecting to developed railway accident ontology. The knowledge graph is a graph-based knowledge derived from large-scale data, aiming to represent real-world entities and their relations from a multi-dimensional perspective (Chen et al., 2020; Liu et al., 2021). In the Semantic Web context, a knowledge graph is constructed for the machine to conduct reasoning and inference over present knowledge to answer queries (Kejriwal, 2019; Kume & Kozaki, 2021). To address the issues of heterogeneous knowledge graphs, the Semantic Web and World Wide Web Consortium (W3C) standards were developed as a universal machine-readable data framework to help the growth and dissemination of knowledge graphs. A knowledge graph consists of a finite set of resource description framework (RDF) triples, containing a subject, a relation and an object. For instance, a piece of knowledge that "Network Rail and Australian Rail Track Corporation are railway infrastructure managers in different countries" can be expressed as two RDF triples: (Network Rail, is_a, railway infrastructure manager) and (Australian Rail Track Corporation, is_a, railway infrastructure manager).

Knowledge graphs can be considered a large ontology with the RDF as a basic unit of stored knowledge and allows the descriptions of elements through the taxonomies of classes and properties. Due to the uniform standards, knowledge graphs are flexible enough to be reused, published, extended and linked to data of interest. Several public and cross-domain knowledge graphs have been published in recent years with semantical structures, including DBpedia, Freebase, OpenCyc, Wikidata and YAGO. All of these knowledge graphs are freely accessible, registered in online dataset catalogues and have well-established practices for interlinking structured data to demonstrate knowledge (Färber et al., 2018). End users are allowed to interact with data in the knowledge graphs or use the programmatical web interface as the initial ontology for further extension. Previous studies have applied this as the foundation of developed knowledge. Metzke et al. (2013) developed a semantic complex event processing for logistics to detect meaningful events (such as a flood) within the transportation route based on the DBpedia top-level ontology. Abdullah et al. (2017) used DBpedia to retrieve the semantic meaning from the recorded voice of control tower operators and pilots to identify aviation safety-related events. Several data-driven ontologies have also borrowed the RDF protocols and constructed knowledge networks from knowledge graphs for NLP tasks, such as entity disambiguation (Alpizar-Chacon & Sosnovsky, 2019; Kume & Kozaki, 2021) and question answering (Sakor et al., 2020).

To connect the data with entities in the knowledge graphs, entity linking is introduced and potential interface issues in the literature are discussed.

3.4.3 Entity linking

To disambiguate heterogenous railway terminology and link entities to the ontology using knowledge graphs, the sub tasks of entity linking are introduced: mention detection, candidate generation and entity disambiguation.

Mention detection (MD) is a task to identify potential entity mentions in the data (Broscheit, 2020), which is also an essential step prior to a variety of NLP downstream tasks such as name entity recognition and co-reference resolution (Yu et al., 2019). The mentions refer to any possible representation of entities, including gazetteer (a geographical dictionary), people's names and co-reference (Lata et al., 2022). Other mentions might direct to the same entity; for instance, level crossing and grade crossing indicate the same concept.

Candidate generation (CG) refers to the task of retrieving entity candidates for each mention from knowledge graphs (Broscheit, 2020). All potential entities in knowledge graphs are reviewed and candidates for linking to mentions extracted from the previous step are recognised. One word with more than one meaning might obtain multiple outcomes from CG. For example, using the word "train" in CG might result in two entities as candidates: the rolling stock, and the process of improving the level of awareness of an individual.

Last, entity disambiguation (ED) is a task leveraging contextual information of data to determine the linked entity to a mention (Broscheit, 2020). Several types of semantic and syntactic clues are applied to score extracted candidates and address mention ambiguity. For example, considering the sentence "The train on platform 3 goes to the Blue Mountains and is departing at 16:30", words surrounding the mention "train" mostly link to the facility of rail transport, gazetteer and time. Therefore, the model for entity disambiguation is expected to assign this mention to the entity "rolling stock".

3.4.4 Evaluation of knowledge graph selection and entity linking

Despite the increasing trend of developing professional domain-knowledge graphs with uniform standards in many fields, the issues of the quality of updates, maintenance of knowledge graphs and entity linking tools have been noted. For instance, Labusch and Neudecker (2020) published a powerful BERT-based repository for entity linking and demonstrated high-quality knowledge graphs but these are currently unavailable due to the lack of maintenance and an out-of-date programming environment. To address these issues, a considerable number of studies evaluate the quality of knowledge graphs by designing a series of metrics from various dimensions (Chen et al., 2020; Issa et al., 2021). In practice, Färber et al. (2018) have summarised a set of operational criteria based on the needs of knowledge graph users, including intrinsic data quality (Piscopo & Simperl, 2019; Shenoy et al., 2022), contextual data quality (Piscopo & Simperl, 2019; Zhang et al., 2019), representational data quality (Frey et al., 2019; Piscopo & Simperl, 2019), and accessibility (Li et al., 2018; Piscopo & Simperl, 2019). Each dimension reflects different considerations of the quality, and trade-off between metrics is inevitable while selecting knowledge graphs due to the conflict between dimensions. For instance, one knowledge graph meeting the requirement of "the knowledge graph covers a basic population of general knowledge" might inevitably violate the criteria "RDFs in the knowledge graph do not contain empty nodes" due to the possible absence of linked real-world data.

To ensure that selected knowledge graphs meet needs of interests, the following seven requirements are included: high consistency for processing big data, only limited domains of interests, high quality of the RDFs' structure, capability of comparison with other knowledge graphs, a low number of absent entities, capability to handle intensive enquiries, and inclusive of temporal information. The evaluation process derived by Färber et al. (2018) is adopted. More details are presented in Chapter 4.

3.4.5 Application of the ontology in the context of railway safety

A considerable number of ontologies in the railway context have been published and used by industry and academia. An overview of existing railway-domain ontologies can be found in Table 3-7 (industry sources) and Table 3-8 (academic sources). Most proposed industry ontologies address specific topics, such as semantic rail data integration (IT2Rail), standards for sharing infrastructure data across organisations (RailTopoModel), and common languages exchange of file formats (IFC Rail). On the other hand, ontologies developed by academic researchers mainly emphasise analysis of operational risks and interface between data resources.

However, most ontologies released from academic studies are one-time based and not updated after publication. Original repository and applied protocols are not available due to the lack of use and maintenance. For the industry ontologies, the maintenance may stop after the end of projects unless there is private support from organisations. Most ontologies for the purpose of data integration across projects stop being updated because of the disappearance of demand, such as InteGRail for offering a standard for data interchange across European Union projects, and SMART-RAIL for exchanging rail freight knowledge with other projects under SHIFT2Rail. Limited industry ontologies are active with the aim of sharing and standardising railway infrastructure data (RailTopoModel and RailML) and signalling data (EULYNX) across European countries. Nevertheless, inconsistent RDF formats and manual rules are applied to reviewed railway domain ontologies, resulting in another obstacle for further reuse. An example is the ontology for railway operational accidents which manually extracts the knowledge entities and their relations in accordance with the check lists and standards proposed by Liu et al. (2021). Despite a comprehensive data connection between RAIB reports ranging from 2005 to 2015 and the ontology, further supplementary data has not able to be implemented due to limited revealed methods and customised RDF standards.

Table 3-7: Rail domain ontologies – Industry sources

Domain	**Project name**	**Description**	**Year of latest update**
Data Integration Ontologies	InteGRail	Offering a standard for data interchange within Europe via Network Statement Checker Ontology (NSO)	2008
	IT2Rail (part of SHIFT2Rail)	Providing IT solutions for semantic rail data integration	2018
	SMART-RAIL (part of SHIFT2Rail)	Providing integrated knowledge to improve rail freight services	2018
	RaCoOn (part of Capacity4Rail)	Demonstrating the ontology-based data integration considering various European sub-systems	2017
	ST4RT	Improving the interoperability based on Shift2Rail project	2018
Standards of rail data formats	RailTopoModel (RTM)	Standards for sharing of infrastructure data across organisations, established and maintained by Union Internationale des Chemins de Fer (UIC)	Now
	RailML	Offering the XML-based standard for railway infrastructure data exchange based on RTM	Now
	EULYNX	Offering standardised signalling systems interface based on RTM	Now
	IFC Rail	Offering a common language to exchange file format with rail-specific parameters	2021
Others	Open Rail Data	Open data from the rail industry in Great Britain, maintained by Network Rail and National Rail	2019

Table 3-8: Rail domain ontologies – Academic sources

Ontology domain	**Description**	**Reference**
Heavy haul railway domain risk ontology	Classifying factors and establishing the top-level risk ontology	(Cao et al., 2019)
Railway data integration ontology	Developing seamless integration of railway data	(Lewis, 2015)
Common safety ontology for railway vehicles	Building top-level railway safety ontology based on ISO 26262, EN 50126 and SIRF (the German Standard for railway vehicles)	(Hulin et al., 2016)
Railway accident scenario ontology	Establishing operational railway accident ontology from the perspectives of error in system, hardware, software and human	(Maalel et al., 2012b)
Railway derailment ontology	Developing the derailment ontology based on case studies	(Zhao et al., 2022)
Multi-lingual railway accident ontology	Integrating railway accident data written in different languages	(Hughes et al., 2019)
Ontology for dysfunctional railway analysis	Building ontology for railway dysfunctional analysis based on Unified Foundational Ontology (UFO)	(Debbech et al., 2020)
Rail topology ontology	Constructing rail infrastructure ontology via XML-based data formats and UML-based object-oriented models	(Bischof & Schenner, 2021)
Rail-road incidents ontology creation and upgrade	Building ontology via data-driven machine learning approaches and structured data	(Pramanik et al., 2021)

3.5 Scoping workshops and surveys as tools for outcome evaluation

This section outlines the qualitative approach adopted involving scoping workshops and surveys to identify the scope of research and evaluate the outcomes of models developed in previous sections. Workshops and surveys have been widely employed in the literature as dual mechanisms to gather comprehensive feedback from both technical and end-user perspectives (Beamer et al., 2007; Roberts et al., 2013; Worton, 2012). This approach acts as an effective way to bridge the gap between research and practice and develops the needs of practitioners.

3.5.1 The workshop and survey design

To help validate the approach proposed and ensure the developed model meets the railway industry's needs, a scoping workshop with interactive sessions and follow-up survey was selected to facilitate in-depth discussions. The workshop includes a showcase containing the outcome of the model designed in this study with the objectives of both the workshop and survey aligned with the models' outcomes and relevant literature. The workshop is designed as a 4-hour event, expecting to cover three aspects: finding dissemination, roundtable discussion, and take-out survey. The finding dissemination section concentrates on sharing the findings of the models. Exercises aiming to observe how participants interpret and understand the models are also included. Participants are given keywords and representative documents and asked to name the topics. They are encouraged to elaborate on how they interpret the topics and relationship with their day-to-day operation. The roundtable discussion focuses on feedback and interpretation. Participants are encouraged to connect the models with practical operation and identify potential contributions of the models as well as any issues.

The participant's survey is designed based on the outcomes of models and theoretical underpinnings of organisational learning framework in the literature (discussed further in Section 7.2). The aims of the survey are to validate the outputs from the workshop and offer opportunities to provide individual experience and feedback. Key themes are identified from the review of the literature and relevant topics identified from topic models, covering the following topics: background information, information receiving, information processing and information disseminating. Questions in the surveys are open-ended addressing possibilities, barriers and future directions. The outlines of the participants' survey are illustrated in Appendix A (Section 9.1). A pilot survey was conducted and distributed to academics for checking the appropriateness (discussed further in Section 4.2).

3.5.2 Sampling strategies

For the expected characteristics of participants, no specific characteristic was required other than at least 10 years of experience in the railway industry. The relevant fields of the railway industry include infrastructure provider/owner, infrastructure maintainer/contractor, operators – managers and workers, regulatory authority (government, railway safety board, accident investigator), and consultant (third-party, academia). The expected number of participants in each workshop was 10 to 15 but not limited. However, the threshold of minimal work experience could be lowered if the number of participants is far lower than expected. Participants were invited to select the preferred time zone for their participation.

The selective sampling strategy was applied to curate a list of workshop participants. Invitations are sent to members of major transportation community mailing lists, comprising professionals engaged with transportation technology, policy, and innovation. The mailing lists include, but were not limited to, the Universities' Transport Study Group (UTSG), a professional discussion forum for public transport researchers and practitioners, and the Australasian Transport Research Forum (ATRF), the principal transport research forum in

Australasia. Several private invitations were also sent to experts in specific contexts, such as railway safety and natural language processing.

3.5.3 Analysis of workshop and survey outputs

The qualitative data generated from the workshop contains two parts: the observation of exercises and the roundtable discussion. The observation content is used to extract the attitude of participants toward the usefulness of models and demands for practical implementation of participants. Such data will become critical inputs when refining the models. A series of labels representing topics are also expected to be extracted by the researcher and used in the roundtable discussion and survey. On the other hand, the roundtable discussion mainly covers the usefulness of the models and underlying issues in topics. Subsequently, the content is coded with an in-depth reading of the transcribed data. The codes are closely aligned with the opportunity, challenges and insightful findings for practical use. The codes reflect the underlying perspectives of participants, identify existing issues and reveal the potential solutions. Finally, these codes are integrated with the conceptual frameworks proposed (see Section 7.2) to mitigate the gap between the practice and academia.

The survey primarily focuses on how the information is received, processed and disseminated by the railway industry across countries. As questions are open-ended and involve more individual experience, the survey is analysed by manually reading, extracting codes used in the workshop and classifying them into existing conceptual frameworks proposed. After this, participants' perspectives and understandings of information reception, processing, and dissemination are reviewed and compared with the outcomes of workshops and the quantitative outcomes of models. This combination ensures a consistent and holistic view.

3.6 Synthesis of findings

This chapter has provided a detailed introduction to the mechanisms of natural language processing models and various training approaches. A comprehensive discussion of the concept of word embedding for textual data and the state-of-the-art language models synthesised the availability of resources and potential limitations. Comparing training approaches derived from the review of the literature helped determine the appropriate method to address the restrictions of the data. The benefits and drawbacks of the rule-based, (semi-) supervised learning based and unsupervised learning based approaches were revealed to address the obstacle of practical implementation.

In addition, the concept of ontology has been discussed. The present railway-related ontologies and the process for developing an ontology have been investigated to understand the reusability of existing ontologies. Despite a number of accident-based and risk-based ontologies in the academic literature (Cao et al., 2019; Debbech et al., 2020; Hughes et al., 2019), the findings suggest that there is no ontology to describe the interfaces between entities and organisations involved in railway accidents that can be reused and extended the research in thesis. Last, the concept of knowledge graphs and their application in the literature and practice has been demonstrated. The entity linking process was also introduced as the solution of the interface between knowledge graphs and real-world data. Additionally, the process for evaluating existing knowledge graphs was discussed to ensure the quality of the developed ontology.

To address the identified methodological issues and answer the research questions, the next chapter elaborates on strategies developed and the process of building models for the automation of railway accident analysis.

4. The development of models for the automation of railway accident analysis

This chapter introduces a series of models for the automation of railway accident analysis. This chapter expands on the discussion of the previous chapters to show how the research objectives and research questions of the thesis are addressed. The structure of the chapter is as follows: the framework of model with topic modelling and entity linking strategy is discussed (Section 4.1), followed by the evaluation of model using scoping workshop and survey (Section 4.2). The synthesis of findings in this chapter is presented (Section 4.3).

4.1 Framework of model

This section shows details of the proposed analysis flow, consisting of the topic modelling, entity linking strategy, covariate analysis, temporal analysis, and the evaluation of developed models. Railway accident reports from the four different countries (UK, USA, Canada and Australia, as introduced in Section 2.3) are first put into topic modelling to extract potential hazards or risks. Subsequently, the heterogeneous terminology used in different areas is standardised with the proposed entity linking strategy. Covariate analysis and temporal analysis allow us to understand how the identified hazards and risks have the potential to influence the railway industry across countries. Last, the developed model is evaluated and reviewed by a workshop and survey.

4.1.1 Topic modelling

Topic modelling is a practical application in information retrieval and the NLP to categorise text into domain topics and rank documents over topics (Dornick et al., 2021; Roque et al., 2019; Yang & Anwar, 2016). A topic model reveals the relationship between topics and documents by different features, such as the probability of occurrence of words and high dimensional word embeddings (see Section 3.3.1 and 3.3.2). The model assumes that a document contains a collection of underlying themes, and the distribution of words in the document over the whole corpus might derive topics representing these underlying themes. A set of keywords is identified to reflect underlying topics and their trend, which is informative statistics for further methodological and practical applications (Blei & Mcauliffe, 2007).

A topic model can be trained in several ways, including supervised learning, semi-supervised learning and unsupervised learning. To ensure high automated analysis and avoid human intervention, unsupervised learning approaches are selected to build the topic model. A considerable number of off-the-shelf programming packages for advanced NLP applications are developed and publicly available, such as Spacy for deep learning workflows and pre-trained language models (Choi et al., 2015; Jugran et al., 2021), StanfordNLP for toolkits used in developing extendable pipeline and pre-trained models (Manning et al., 2014), and NLTK for a wide range of libraries for statistical NLP and data preprocessing purposes (Bird & Loper, 2004). Several package-oriented programming models are developed based on these packages and the state-of-the-art technologies introduced in Chapter 3 and result in significant improvements of performance in the topic modelling contexts, such as the structural topic model (STM) (Kwayu et al., 2021; Li et al., 2011; Roberts et al., 2019) and BERTopic model (Grootendorst, 2022). The following sections elaborate on the details and applications of the STM and BERTopic models.

Structural topic model

The structural topic model (STM) is an unsupervised learning based probabilistic topic modelling method derived from the latent Dirichlet allocation (LDA). The LDA is a generative statistical model that classifies documents based on the observations of each individual word collected in the documents and assumes that the topic of each document is derived from the aggregation of the words in that document. Suppose a word is a basic item from a set of vocabulary indexed by {1, 2,…, V}, a document (w) is a sequence of N words noted by w =

($w_1, w_2, \dots, w_N$), and a corpus is a set of M documents noted by $D = \{w_1, w_2, \dots, w_M\}$. Assume documents ($D$) are created by a random combination of latent topics, characterised by a specific distribution over words (N) and follow a generative probabilistic model (Blei et al., 2003). The generation of each document (d_i) in a corpus D follows the consecutive theorems:

(1.) The number of words N is chosen by a Poisson (ζ) distribution.

(2.) A random parameter θ drawn from a Dirichlet (α) distribution is chosen to represent the proportions of topics in one document.

(3.) For each word w_n in N words within one document, a random topic z_n is assigned to w_n drawn from a multinomial (θ) distribution.

(4.) For each topic z_n, proportions of each word are drawn from another multinomial distribution $p(w_n|z_n, \beta)$, where β is a parameter representing the proportions of words in one topic.

The basic LDA model is commonly applied and virtualised as explained by Blei (2012) and shown in Figure 4-1. Assuming that the dimensionality of the Dirichlet distribution is a fixed and known value k representing the number of topics, the β can be parameterised as a $k \times V$ matrix for mapping the probabilities of words based on the bag-of-words approach. Blei (2012) also notes that N is an independent variable, and its randomness is ignored during the development of the LDA model. Thus, the probability density of the proportions of topics in one document retrieved from the Dirichlet (α) distribution can be illustrated as:

$$p(\theta|\alpha) = \frac{\Gamma(\sum_{i=1}^{k} \alpha_i)}{\prod_{i=1}^{k} \Gamma(\alpha_i)} \times \theta_1^{\alpha_1 - 1} \cdots \theta_k^{\alpha_k - 1} \qquad \text{Equation 4-1}$$

where $\Gamma(\mathrm{x})$ is the Gamma function and α_i is a k-vector mapping the distribution of topics. Equation 4-1 reflects two “plates” in Figure 4-1, representing documents (M) and the recurring choice of words and topics in one document. The outer plate represents the association between all documents (M) and the random parameter θ drawn from a Dirichlet (α) distribution, whereas the inner plate illustrates the association between all words (*N*) in one document and random topics z_n drawn from a multinomial (θ) distribution. Therefore, the link between topics and words appearing in each document is built. On the other hand, another parameter β is estimated to identify the link between the proportions of words in one topic. The joint distribution of θ for a set of words *w* and topics *z* can be expressed as:

$$p(\theta, z, w|\alpha, \beta) = p(\theta|\alpha) \times \prod_{n=1}^{N} p(z_n|\theta)p(w_n|z_n, \beta) \qquad \text{Equation 4-2}$$

In this case, key parameters α and β can be inferred with a Bayesian approach by estimating the posterior distribution of known variables from the original corpus (Kuhn, 2018).

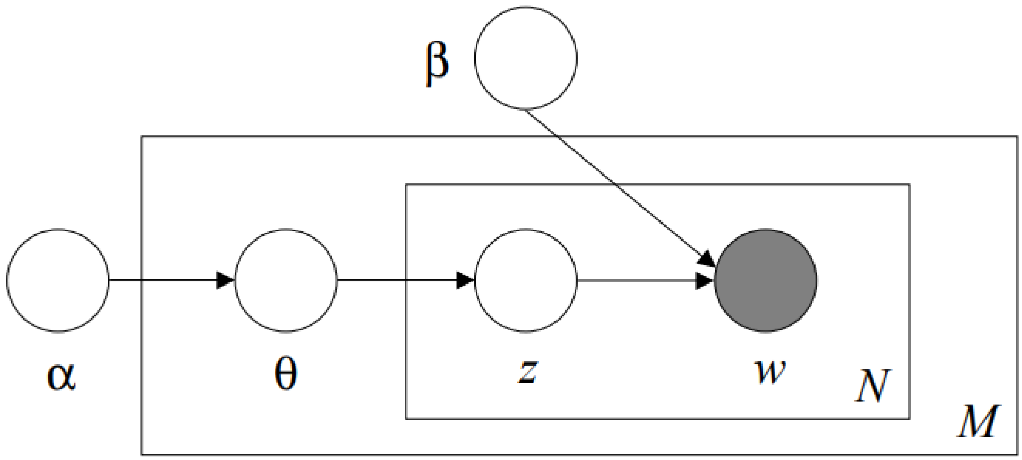


Figure 4-1: The concept of the LDA model illustrated as a "plate" diagram (Blei et al., 2003)

The LDA has been widely improved and implemented in accordance with the context of interests. For example, Li et al. (2018) advanced the structure of the LDA model by training a word2vec embedding on the dataset. The journal articles dataset was partitioned into summary, method and conclusion according to the cosine similarity of embeddings. A weighted topic embedding is created to improve the accuracy of the clustering result. Another example is that Guo et al. (2019) improved the accuracy of LDA by partitioning the documents into paragraphs and applying weighted summation to obtain the predicted topics. Despite the convenience of retrieving document-level information delivered by the LDA, the assumption that the probability of the occurrence of one word within one document is fixed after the LDA model is developed restricts the flexibility of analysis. For instance, estimated parameters α and β are not allowed to be sensitive to temporal factors or other potential covariates (Kuhn, 2018).

The STM is developed on the same statistical basis as the LDA in addition to allowing correlations of external factors among topics. The main difference lies in the pre-generalised linear models derived from the nature of the data used while estimating parameters. In doing so, the parameter θ is not applied to all documents equally drawn from the Dirichlet (α) but from the logistic-normal distribution to estimate the topical prevalence on document-level data. Furthermore, the assumption is that fixed parameter β (distribution of topics over words) should be addressed by replacing the multinomial distribution with a multinomial logit model for estimation. Mathematically, the parameter ($\beta_{d,k,v}$) for an individual word *v* in document *d* within the topic *k* should be as the following equation for capturing the influence of covariate data (Roberts et al., 2013):

$$\beta_{d,k,v} \propto \exp\left(m_v + \varphi_v^{\cdot,k} + \varphi_v^{y,\cdot} + \varphi_v^{y,k}\right) \qquad \text{Equation 4-3}$$

where m_v is the baseline occurrence of word *v*, the $\varphi_v^{\cdot,k}$ is the effect of topic *k*, the $\varphi_v^{y,\cdot}$ is the effect of covariate *y*, and the $\varphi_v^{y,k}$ is the mixed effect among topic *k* and covariate *y*. Thus, the plate diagram (Figure 4-1) can be further extended in Figure 4-2. The main distinction lies in the prior estimation of parameter θ during topic prevalence analysis and additional consideration of covariate variables in topical content. More mathematical details and theorems can be found in Roberts et al. (2013). The STM is more suitable than the LDA for analysis of railway accident reports because critical covariates are usually disclosed and discussed in reports, such as the occurrence of time and the involved mode of rail transport and organisations. These critical covariates can offer valuable insights for better understanding the nature and prevalence of railway accidents across time. For instance, the

STM may provide the difference in how platform–train interface incidents occur on a light rail system and other modes of rail transport. The trend of how it happens may also be revealed by supplementing the occurrence of time as an additional covariate in STM temporal analysis.

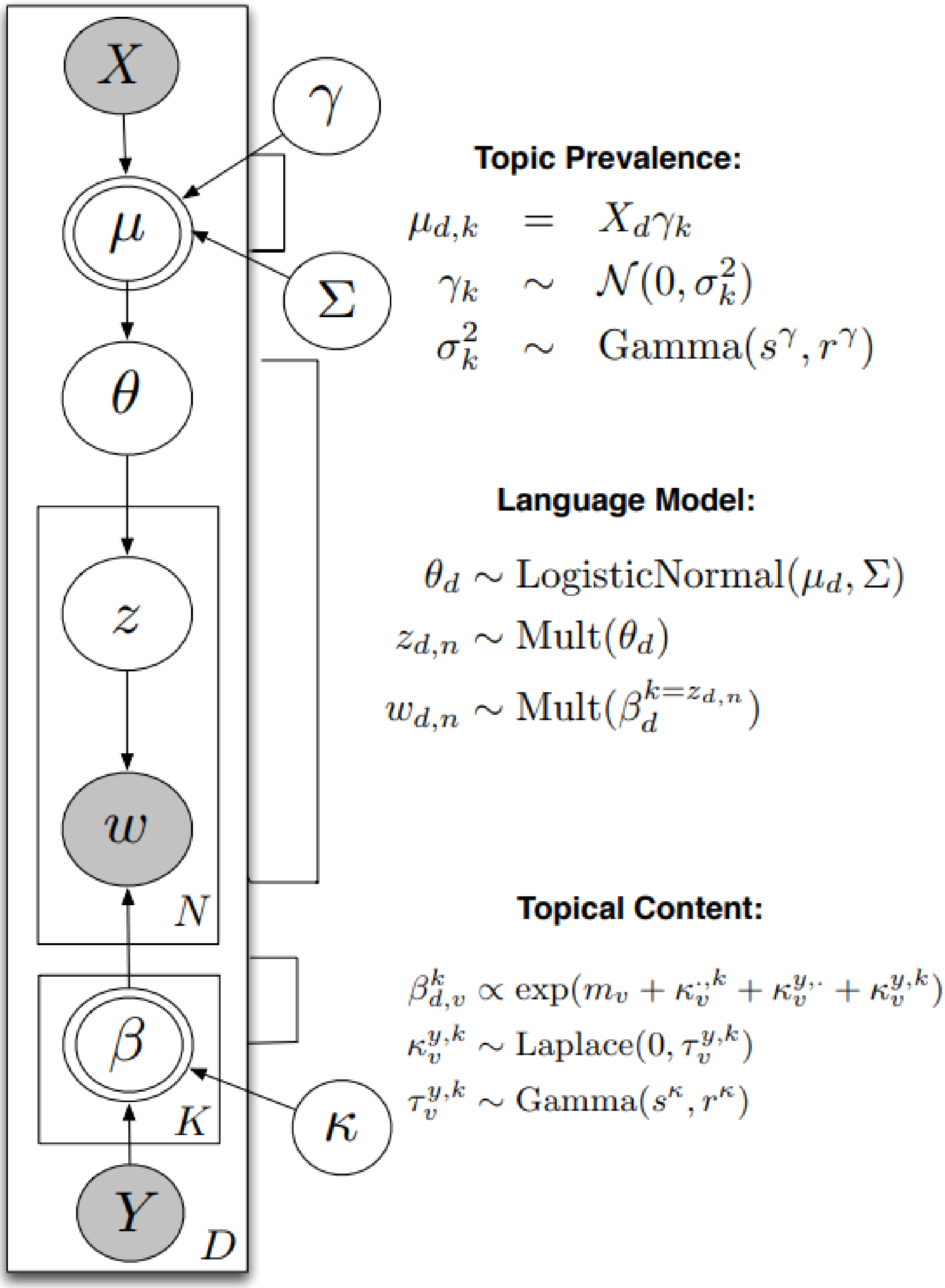


Figure 4-2: The concept of the structure topic model illustrated as a “plate” diagram (Roberts et al., 2013)

To ensure the performance of the developed model, two metrics are introduced as indicators: semantic coherence and exclusivity. The semantic coherence is a measurement determining the occurrence of individual words and the co-occurrence of the pairs of distinctive words. For instance, the thesis dataset has terms “freight train” and “passenger train” with the same word “train”, and a distinct topic should be able to detect this and assign these two words to different topics. On the other hand, exclusivity means the extent to which the model is able to assign one critical keyword to one topic with a high level of appearing possibility and ensure the possibility of appearing is low in other topics. A higher semantic coherence usually leads to lower exclusivity and *vice versa*. A model with a lower number of topics (*k*) would have higher semantic coherence because a limited number of topics and words are used for estimation. However, it would lead to lower exclusivity as well because the option of critical keywords is limited. When the *k* increases, the semantic coherence would decrease, whereas the exclusivity increases because more topics are available for assigning distinct keywords. Once *k* is equal to the number of words in the vocabulary (*V*), the exclusivity will become almost infinite, and the result will not offer any valuable insight. To reach a balance between semantic coherence and exclusivity by determining a suitable number of topic *k*, Equation 4-4 covered by both metrics is designed in this thesis to estimate the balanced performance of the developed STM.

$$Performance_i = \left(\frac{SC_i - SC_{max}}{SC_{min} - SC_{max}}\right) \times \left(\frac{\text{Exclusivity}_i - \text{Exclusivity}_{max}}{\text{Exclusivity}_{min} - \text{Exclusivity}_{max}}\right) \qquad \text{Equation 4-4}$$

where SC_i represents the semantic coherence of the *i* model, SC_{max} and SC_{min} are the highest and lowest semantic coherence values in all models respectively, Exclusivity_i represents the exclusivity of the *i* model, Exclusivity_{max} and Exclusivity_{min} are the highest and lowest exclusivity values in all models respectively, allowing us to select the model with the lowest marginal effect on semantic coherence and exclusivity for a range of *k*.

BERTopic model

Although the STM provides additional functions for understanding underlying themes and trends of a corpus, the foundation of the language model used in the development of STM is based on bag-of-words representations. The bag-of-words approach treats each word individually regardless of the relationships between words, hindering the model from taking account of the context and semantic information in the text. To address this issue, a growing number of researchers use the word embedding representation to accurately capture the contextual information of textual data in the NLP field (Dang et al., 2020; Han & Eisenstein, 2019; Heidarysafa et al., 2019). The semantic information can be adequately captured by vectorising texts.

The BERTopic is a topic model adopting the BERT (see Section 3.3.1.3) pre-trained language model (Devlin et al., 2018) to retrieve high-dimension vectors of texts for clustering. For implementation, topics are generated through three steps: text vectorisation with a pre-trained language model, dimension reduction for optimising the modelling process, and topic representations with custom class-based TF-IDF (c-TF-IDF).

For the text vectorisation with a pre-trained language model, documents in the corpus are embedded in vector space with high dimensions, allowing semantical comparisons. For instance, the sentences “The train stops before the signal” and “The train fails to stop before the signal” will have a longer semantical distance in vector space than the representation created by the bag-of-words approach. The Sentence-BERT (SBERT) framework (Reimers & Gurevych, 2019) is used to convert texts into dense vector representations, which has been commonly applied to NLP tasks and achieved high performance (Ganesh et al., 2020; Labusch & Neudecker, 2020; Ly et al., 2020). The author of BERTopic also states that the language model used in BERTopic is exchangeable so that the performance can be continuously improved through the development of NLP techniques (Devlin et al., 2018).

Once the dense vectors are generated, the spatial distance between data becomes less meaningful due to the multidimensions of local and global features. Therefore, the Uniform Manifold Approximation and Projection (UMAP) technique is introduced to reduce the dimensionality by projecting vectors to lower dimensional space (McInnes et al., 2018). Subsequently, the Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN) is used to cluster vectors in lower dimensional space (McInnes et al., 2017). The advantage of HDBSCAN is allowing noise to be modelled as outliers, avoiding unrelated documents being sorted to topics and influencing the representations of topics. The clustering approach can be replaced by other algorithms in the interest of accuracy and computational time.

Last, each identified cluster is assigned to one topic with a distribution of keywords. To highlight the difference between clusters, the custom class-based TF-IDF (c-TF-IDF) is used to rank keywords by the combination of term frequency (TF) and inverse document frequency (IDF) (Devlin et al., 2018; Hakim et al., 2014). The weight of a term (*t*) over documents sorted to a topic (*c*) can be expressed as Equation 4-5:

$$W_{t,c} = tf_{t,c} \times \log\left(1 + \frac{A}{tf_t}\right) \qquad \text{Equation 4-5}$$

where *tf* is the term frequency and *A* is the average number of keywords per topic. The output reflects the importance of a term in one topic rather than in one document, allowing us to understand the distributions of keywords on each topic. Furthermore, Equation 4-5 can be extended for dynamic topic modelling to reflect the evolution of topics over time. For instance, a topic relating to “over speeding” and “SPAD” can be found across the corpus, but the term “Eurotunnel” might not be found in documents before 1994 (the date of opening). Such variance has been mixed, hindering researchers from understanding the temporal effect of “over speeding” and “SPAD” on the term “Eurotunnel”. To overcome such difficulties, Devlin et al. (2018) modified the calculation of the weight $W_{t,c}$ by creating a local temporal representation at time *i* with the original equation.

$$W_{t,c,i} = tf_{t,c,i} \times \log\left(1 + \frac{A}{tf_t}\right) \qquad \text{Equation 4-6}$$

Equation 4-6 adds an additional dimension to the weight and enables the representation of local variables without modifying the parameters of the trained BERTopic model and clustered documents. An overview of processes for establishing a BERTopic model is illustrated in Figure 4-3.

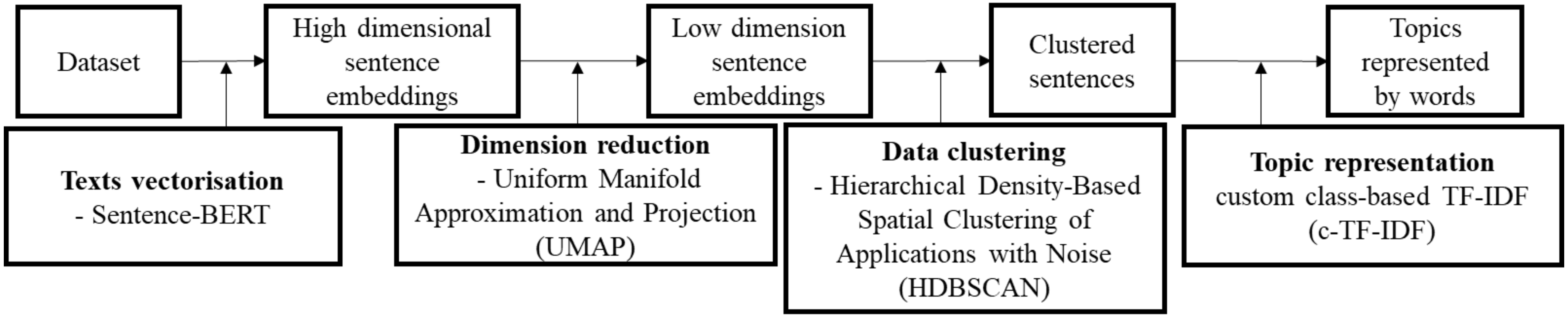


Figure 4-3: Overview of workflows for developing a BERTopic model

Unlike the STM, the parameter required to be set by users in the BERTopic model is not the number of topics but the minimum cluster size (MCS). The minimum cluster size determines the extent to which the HDBSCAN condenses the cluster hierarchy while connecting words for potential topics, which directly influences the number of outliers and the number of identified topics. Figure 4-4 shows the distribution of the number of documents over each topic. The number of identified outliers (which are assigned as topic -1[3] and drawn on the left of the red line) almost dominates 50% of all documents. On the other hand, most topics (after topic 20) contain a very limited number of documents. Such an imbalanced result suggests that either an inappropriate minimum cluster size is set, or an inappropriate dataset is used.

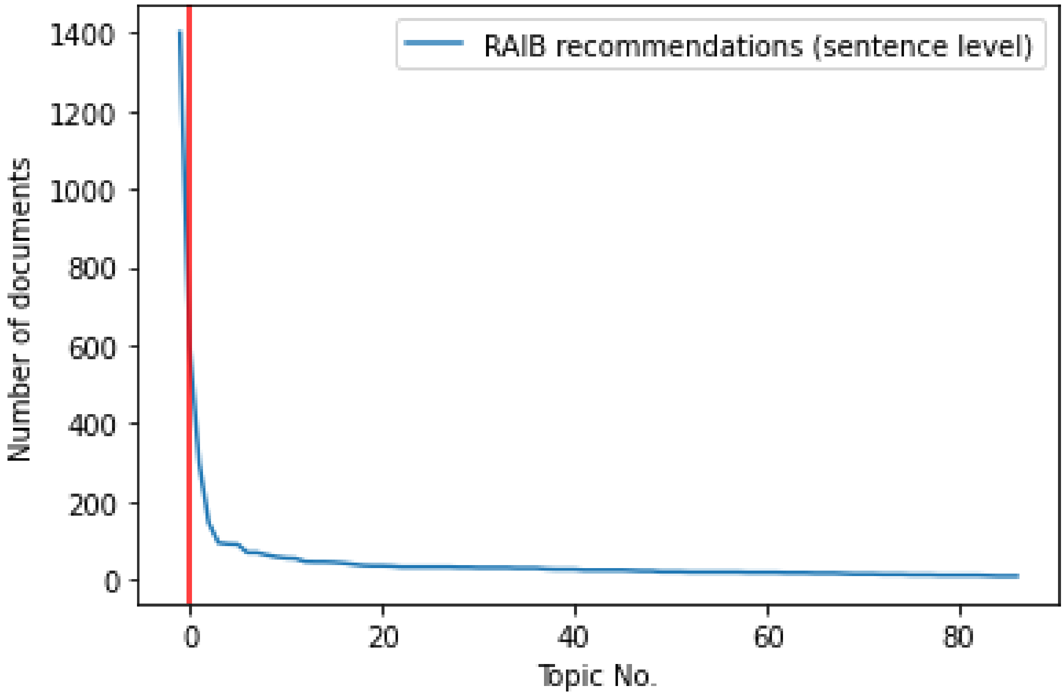


Figure 4-4: Distribution of the number of documents over each topic using BERTopic on the recommendations proposed by RAIB reports (see Section 5.1 and 5.2 for more details of RAIB dataset)

[3] In python programming, the index -1 refers to the last items of a list. Therefore, outliers are distributed to topic -1 which is the last group of a topic list.

It has not been commonly agreed how to determine the minimum cluster size to retrieve a balanced outcome. Devlin et al. (2018) argued that setting minimum cluster size as 15 can lead to better performance for general data analysis. In this thesis, a variety of minimum cluster size values are tested for each dataset used. Outcomes of models with a lower number of outliers and smoother distribution of documents over all topics are manually selected.

Comparison and selection of applied models

Despite a growing body of evidence in the literature showing that word-embedding-based NLP models outperform traditional statistics-based models (Angelov, 2020; Grootendorst, 2022; Han & Eisenstein, 2019; Lata et al., 2022; Ly et al., 2020), the selection of approaches used for topic modelling context via unsupervised learning still heavily relies on the characteristics of data. From the perspective of clustering methods (Figure 4-5), BERTopic identifies a topic by drawing lines to distinguish words in different topics on the higher dimension. On the other hand, the STM obtains topics by iteratively simulating potential topic locations based on word occurrence in each document. Therefore, BERTopic tends to actively connect datapoints close to each other and assign datapoints far from others as outliers. In contrast, the STM keeps modifying the grouping strategies by moving the “angle” of each topic without considering the distance between datapoints. BERTopic might be suitable for the data retrieved from various sources and concentrating on multiple contexts, while the STM might provide more valuable insights if the heterogeneity of data sources and terminology used is low.

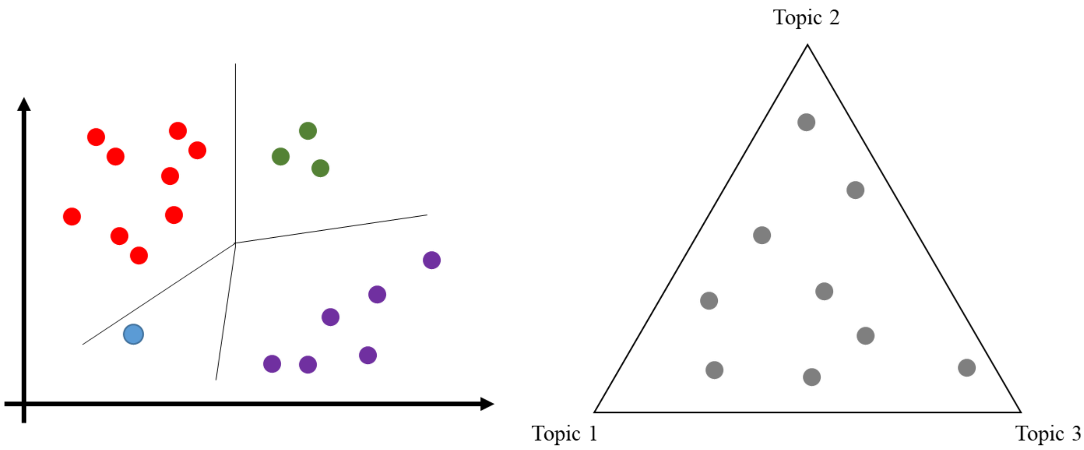


Figure 4-5: Concept of clustering approach used by BERTopic (left) and STM (right)

Another notable difference between the STM and BERTopic is the dimensionality of processed data during clustering. The dimensionality refers to information representing the meaning of words. Higher dimensionality can distinguish words from greater granularity but also increase the computational time during the training process. For the bag-of-words approach used by the STM, the dimensionality is equal to the volume of vocabulary used in the data. Each word is treated individually and uniquely regardless of words with similar meanings. For instance, “rail” and “track” are identified as two independent words in the STM even though they share part of the concept that supports wheels to roll on. On the other hand, the word-embedding-based approach (BERT) identifies the meaning of one word by considering the words in the same document. The higher dimensionality of the word-embedding-based approach allows the model to store the characteristics of each word from different dimensions. BERT, one of the most popular language models, uses 768 dimensions for machines to understand the context of texts (Devlin et al., 2018). For instance, the words “rail” and “track” share the similar distribution of dimensions given that these two terms frequently appear in the same context.

The dimensionality of processed data plays an important role in determining the granularity of input data. In other words, the decision of whether to split documents into sentences before developing analysis needs to be made for the appropriate balance between the level of captured information from texts and computational requirements. An example of dimensionality of 339 processed railway accident reports published by the RAIB containing 13,077 vocabularies and 15,294 sentences is illustrated (Table 4-1).

For the STM, the dimensionality is associated with the volume of vocabulary. Thus, the transferred data matrix is sparse, not sensitive to the length of documents, and contains only “1” and “0” (Pennington et al., 2014). On the other hand, the BERT-based model compresses input data and expresses each datapoint with 768 dimensions regardless of the number of words in one document, implying that long documents might be over-condensed, and much information will be lost during transformation. Although the literature has not suggested the appropriate data length for BERT model transformation, a full railway accident report containing different contextual sections and over 10,000 words might be too rich to be compressed into (the popularly used) 768 variables. A large-size dense matrix also requires heavy multiplications and increases already long computational time; for instance, 260,352 multiplications must be done to develop a BERT-based model with a 339×768 matrix for document-level analysis. Last, the BERTopic model introduces the uniform manifold approximation and projection (UMAP) technique to project higher dimensional features to lower dimensional expressions. Although part of the information is sacrificed, the outcome retrieved from compressed higher dimensional features provides additional characteristics for topic modelling and requires less time for model development (Grootendorst, 2022; McInnes et al., 2018).

Table 4-1: A demonstration of the dimensionality of processed data over different approaches on the railway accident reports published by the RAIB

Data – railway accident reports (RAIB)	STM	BERT only	BERT+UMAP (BERTopic)
Dimension (input) – document level	(339, 13,077)*	(339, 768)	(339, 3)
Dimension (input) – sentence level	(15,294, 13,077)*	(15,294, 768)	(15,294, 3)
Dimension (output) – document level	(339, n)**	(a, b, 768)***	(a, b, 3)***
Dimension (output) – sentence level	(15,294, n)	(a, b, 768)***	(a, b, 3)***

*Sparse matrix, **number of topics, ***a: number of clusters; b: number of clustered documents

To select the appropriate approach, the structure of railway accident reports should be analysed. Generally, railway accident reports consist of abstract, summary, analysis, conclusions and recommendations sections. Each section illustrates the railway accident from different perspectives. For instance, the abstract and summary sections describe the general information, fact of events, sequence of occurrences and consequences, whereas analysis, conclusions and recommendations emphasise the observations, investigations and comments made by investigators. In addition, a mixture of critical information is also outlined, including causal factors, underlying factors, contributing factors, and identified hazards. All of these indicate that a railway accident report contains a wide range of heterogeneous information that might not be fully captured by the document-level analysis. One document may also include more than one topic, so assigning one document to a single topic via document-level analysis might not be realistic. Nevertheless, the terminology used in the huge corpus may not be consistent over time, implying the STM based on the bag-of-words approach might not be applicable. Therefore, BERTopic used for sentence-level topic modelling is selected as the method for the railway accident reports analysis in this thesis.

On the other hand, BERTopic might not be suitable for the recommendation dataset given that a strong semantic homogeneity of descriptions is commonly found, leading BERTopic to extract such features and recognise others as outliers. For example, consider the following recommendations from the RAIB recommendations dataset (see Section 5.3.2 for more details about the RAIB recommendations dataset):

- It **is expected** that **Network Rail** will take account of principles identified by recent research when modifying **crossings**.
- **Network Rail should** review the design of long hoods that can be fitted at **level crossings** and implement any necessary changes identified to make them more effective.
- When addressing risks identified by the implementation of the revised process, **Network Rail should** prioritise the implementation of required mitigation measures to **level crossings** where consequences of operator error are severe and not protected by engineered safeguards.

The recommendations above are assigned to the same topic by BERTopic because the semantic meaning of Network Rail's obligation on level crossing risks is detected by words in bold, including "expected", "should", "Network Rail" and "level crossings". However, the topic of interest in this study is "how" recommendations address the risk. Keywords with underline including "review", "implementation" and "principles" should be identified and assigned to topics. In this case, the STM might be more applicable because the occurrence of words is more meaningful than the semantic context information. Therefore, the STM is applied to the analysis of the recommendations in this thesis.

4.1.2 Entity linking strategy

There are several approaches for standardising identical mentioned entities with different terms, such as named entity recognition (Li et al., 2020; Settles, 2003; Wu et al., 2020) and information retrieval (Balali et al., 2021; Gudivada & Arbabifard, 2018). However, most models in the literature are developed by supervised learning with a large amount of labelled data, complicated training algorithms and powerful machines. No pre-trained model in a railway-specific context has been developed in the literature. To address the issue without training a supervised learning model, this study develops a series of alternative strategies to achieve the objective of standardisation.

Next, the overview of the entity linking strategy was illustrated as depicted in Figure 4-6. The initial railway accident ontology was established by collecting existing ontologies published in the literature or railway industry and manually reviewing railway accident reports. Subsequently, the knowledge graphs were introduced for identifying and disambiguating potential entities from the railway accident report corpus. A framework for selecting the appropriate knowledge graph was borrowed from the literature, and the result shows that Wikidata meets the requirements of this study. The off-the-shelf API toolkit *Tagme* was used to identify potential candidates for the other entities supplemented in the railway accident ontology and provide the interface between the original railway accident reports and Wikidata. An additional context-sensitive disambiguation process based on the graph theory was proposed on the basis of the existing framework to augment the power of the proposed model on entity detection and entity linking. Last, the railway accident ontology linked to real-world data was finalised. A notable limitation of the derived ontology is that an entity that does not exist in Wikidata may not be extracted and found. The primary use of the proposed ontology is to standardise the heterogeneous terminology used in different countries. In addition, the ontology is also expected to bridge the gap between the identified keywords from the topic modelling and the original railway accident reports (Figure 4-7). The output of the topic modelling is in the form of individual words, which are difficult to interpret. However, the associated entities of each keyword can help better understand the true meaning of retrieved keywords by linking keywords back to the original reports.

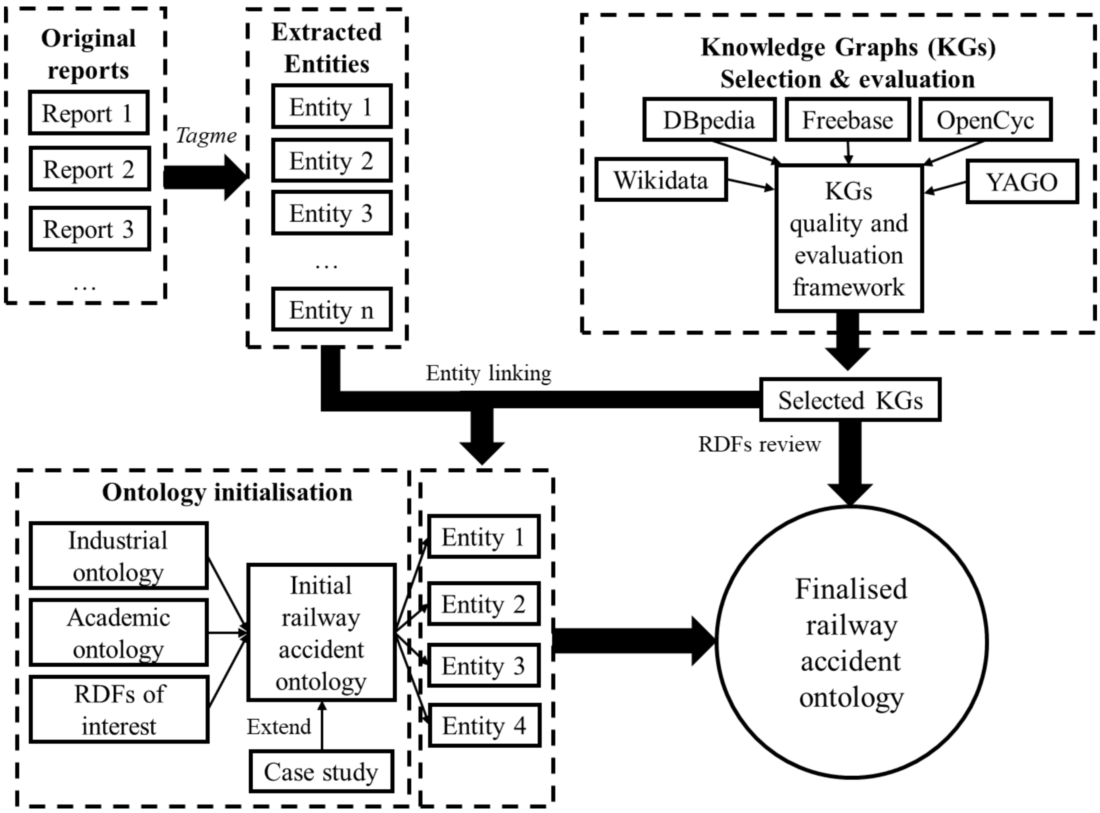


Figure 4-6: Overview of the proposed entity linking strategy

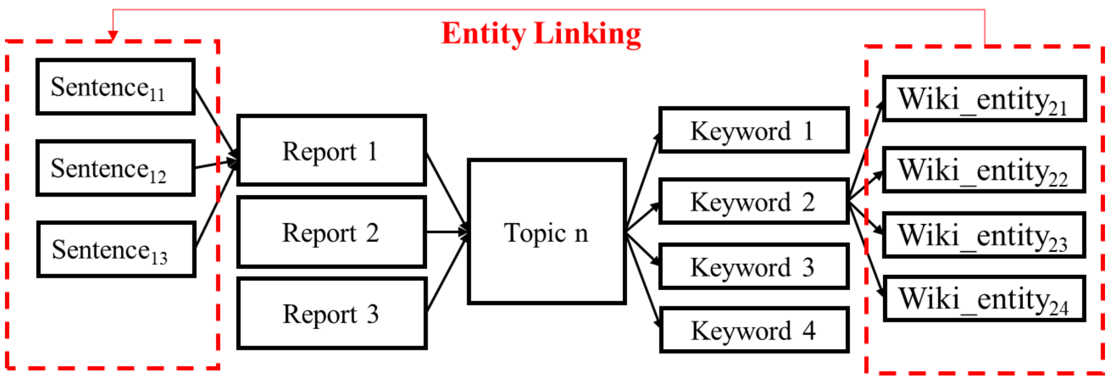


Figure 4-7: Bridge between identified entities and keywords from the topic modelling

4.1.2.1 The proposed conceptual framework for entity linking

The proposed conceptual framework for addressing the entity linking task is illustrated in Figure 4-8. First, an ontology describing the nature of railway accidents, potential underlying factors and systematic interactions between railway organisations is developed (Figure 4-9). Second, a knowledge graph (described in Section 3.4.2) is selected with a comprehensive process proposed in the literature. The knowledge graph is presented as an ontological framework and linked to the designed railway accident ontology to build a hierarchy ontology. Only limited properties are extracted to reduce computational time. Third, an off-the-shelf system *Tagme* is implemented for addressing mention detection (MD) and candidate generation (CG) tasks (Ferragina & Scaiella, 2010). Subsequently, a custom entity disambiguation (ED) strategy is proposed to exclude irrelevant entities. Finally, the remaining entities are linked to the knowledge graph as instances. Therefore, the framework is able to connect terms with the same concepts automatically.

This thesis adopts the evolutive approach to build the railway accident ontology. The objective of this ontology is to provide a map for the entity linking model to connect the terms and entities of interest; outline entities involved and the general mechanism of risks in events; and identify the existing interfaces within the railway industry from the socio-technical perspective. The scale of developed ontology might not cover all concepts but is sufficient for addressing the entity linking task and standardising terminology.

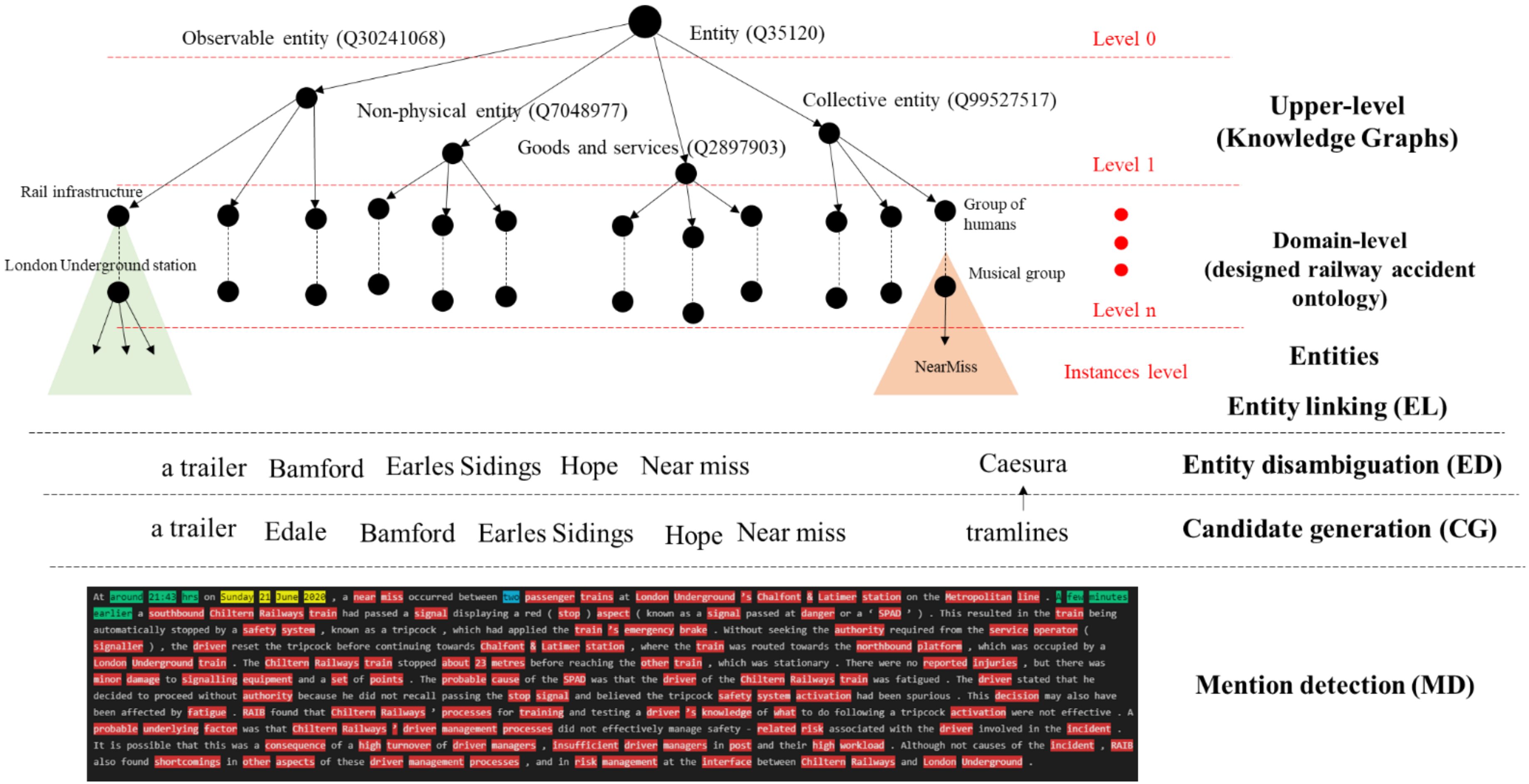


Figure 4-8: Illustration of the proposed conceptual framework to address the entity linking task

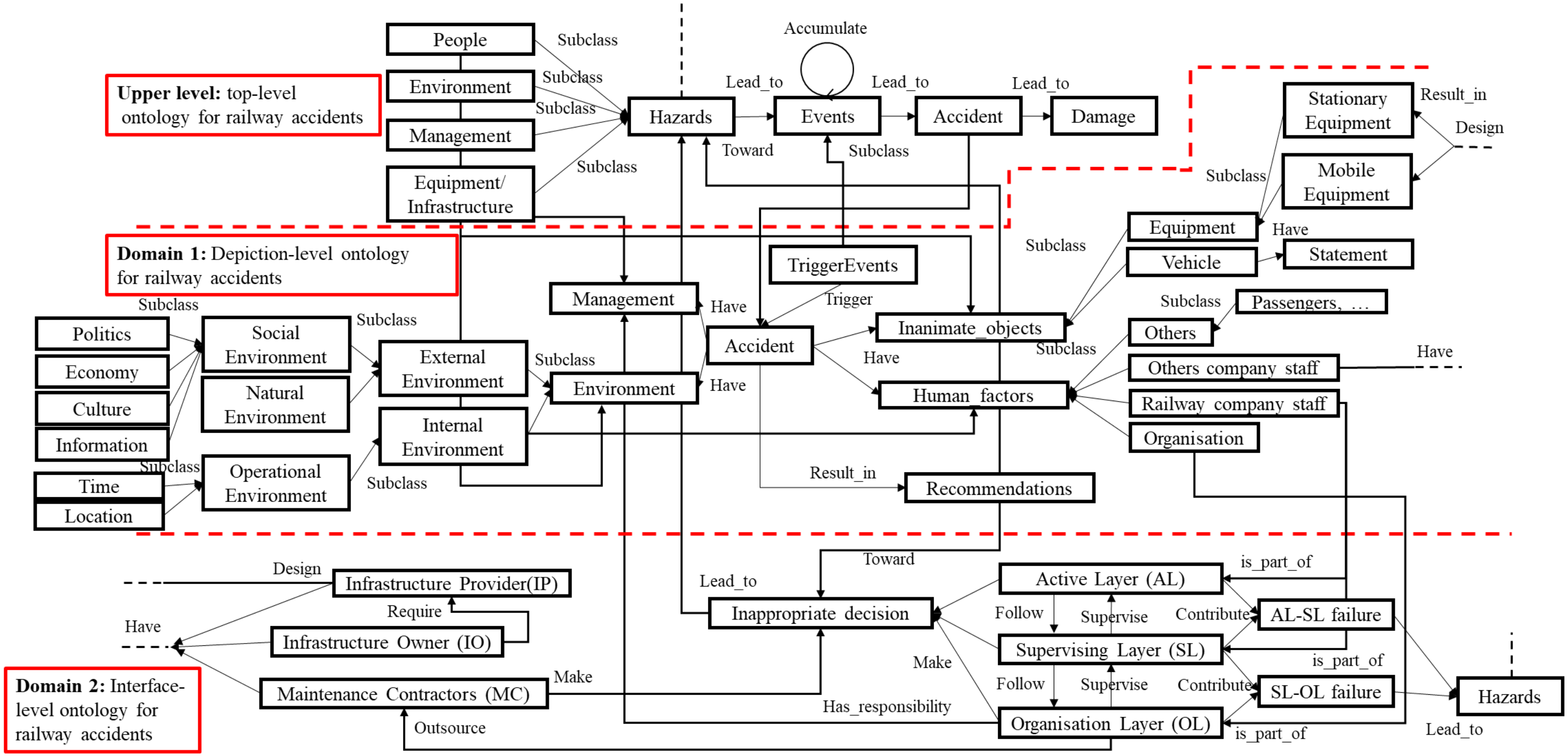


Figure 4-9: The designed railway accident ontology

4.1.2.2 Establishing and refining the railway accident ontology

To start with, an initial railway accident upper-level ontology is established after reviewing existing upper-level ontologies developed in the railway safety literature (Section 3.4.5). Subsequently, ten railway accident reports published by the RAIB are randomly selected and reviewed carefully as the initial map. Critical entities are manually identified and extracted to build the depiction-level ontology. The ontology is further extended with topics identified by developed topic models and the selected knowledge graphs.

Specifically, the process of establishing and refining the ontology is:

(1.) Collect published rail domain ontologies in the literature or the railway industry.

(2.) Review collected ontologies and initialise railway accident ontology.

(3.) Initialise the classes, properties and statements in the ontology.

(4.) Select open-sourced knowledge graphs (KGs) as the interface between the rail accident reports and the ontology.

(5.) Build the metrics for the evaluation of knowledge graphs.

(6.) Use the off-the-shelf *Tagme* package to identify the potential entities (or instances) and their RDFs from railway accident reports.

(7.) Review the RDFs of entities from railway accident reports. Irrelevant entities with classes and properties are removed.

(8.) Identified entities with properties instanceOf (wdt:P31) and subClassOf (wdt:P279) and linked to railway-related domains in the knowledge graphs are extracted and added to existing railway accident ontology after manual reviews (Snowball search) (Kume & Kozaki, 2021).

(9.) Further extend and evaluate the railway accident ontology by context-sensitive disambiguation.

(10.) Finalise the railway accident ontology.

4.1.2.3 Knowledge graph selection and evaluation

Next, a KG needs to be selected to connect the potential entities in texts to the railway accident ontology. Several KGs are publicly available and popularly used in the literature. For instance, Matthew English (2018) developed an extendable schema for constructing annotated corpus with Wikipedia Toolkit API and Stanford Named Entity Recogniser. Wikidata, a collaborative knowledge graph, containing common knowledge of open data, has been applied to address the issue of the homogeneity of terminology used (i.e. rail traffic controller, train dispatcher, train controller, signalman) and ambiguation (i.e., conductor, conductor track). To select the appropriate KGs, the KG quality and evaluation framework proposed by Färber et al. (2018) is applied (Table 4-2). Five KGs used in the literature are identified for evaluation: DBpedia, Freebase, OpenCyc, Wikidata and YAGO (Augenstein et al., 2016; Elsahar et al., 2018; Färber et al., 2018; English, 2018; Sakor et al., 2020).

The evaluation consists of four dimensions: intrinsic, contextual, representational data quality, and accessibility. The intrinsic dimension measures whether the data can be assessed independently from its context. The selected KG should be correct, reliable, free of error, consistent with accumulated knowledge, and accepted to be corrected. On the other hand, the contextual dimension highlights the usability of the KG from the perspective of end users. The selected KG must offer broad and deep knowledge with timeless updates and display in an intuitive order. The representational data quality dimension measures the extent to which data in the KG is human- and machine-readable without ambiguous concepts. Last, the accessibility dimension ensures that the KG is fully open-sourced and linked to other KGs with appropriate interfaces. The selected KG with high accessibility should be available, retrievable and reusable regardless of time and number of requests. Table 4-2 summarises the evaluation framework and describes each criterion.

Despite the completion of the evaluation framework, not all criteria should be treated equally due to the nature of tasks needed during the development of the entity linking model for the railway context. First, entities in the KG should be accurate, cover a wide range of contexts and be connected to other entities with reliable sources. Railway accident reports might contain mixed contexts, such as meteorology, physics, aerodynamics and engineering. The KG is expected to cover terms mentioned in the reports and have the corresponding correct concept describing the identical entity used in original reports. Second, the KG should provide alternative vocabulary under the same entity. The purpose of the entity linking is to standardise the different terminology used in railway accident reports across countries. More alternatives empower the model to identify more heterogeneous terms and potentially increase the performance. Last, the KG should be fully open-sourced with high computer readability and consistency. The KG is used to iterate multiple times on a variety of inquiries for disambiguation purposes. The interface between the KGs and the developed model also must exchange information efficiently for higher programming stability.

After determining the required characteristics, each criterion in the KG quality and evaluation framework is reviewed carefully, and a different level of consideration is assigned to each criterion based on the requirements. The result of evaluating KG candidates is shown in Table 4-3. The score of each criterion is extracted from the original work of Färber et al. (2018). The consideration is valued by this study based on the aforementioned requirements. A criterion labelled with “High” is weighted three times as much as the base score, while “Medium” is weighted two times as much as the base score and a criterion labelled “Low” is treated equally to the original score. The Unweighted Average (*UA*) of the KG *k* is calculated by Equation 4-7.

$$UA(k) = \frac{\sum_{i=1}^{n} m_i(g)}{n} \qquad \text{Equation 4-7}$$

where m_i is the criteria in the KG quality and evaluation framework, *n* is the number of criteria. The Weighted Average (*WA*) can be calculated by Equation 4-8.

$$WA(k) = \frac{\sum_{i=1}^{n} w_i \times m_i(g)}{\sum_{i=1}^{n} w_i} \quad \text{Equation 4-8}$$

where w_i is the consideration indicators. w_i is 3 for the consideration "High", 2 for the consideration "Medium", and 1 for the consideration "Low".

Overall, Wikidata exceeds other KGs on an unweighted and weighted average, indicating that Wikidata generally performs well on each dimension and satisfies requirements. Therefore, Wikidata is selected as the base for the KG for addressing the entity linking model.

Table 4-2: Knowledge graph (KG) quality and evaluation framework (revised based on Färber et al., 2018)

Metrics	Description	Aspects of quality
Intrinsic dimension – Accuracy		
Syntactic validity of RDF documents	RDFs in the KGs are created via standardised tools.	Consistency of data type
Syntactic validity of literals	Values stored are syntactically valid and consistent.	Consistency of data type
Semantic validity of triples	The statements of entities are held true or from trusted sources.	Trustworthiness of data
Intrinsic dimension – Trustworthiness		
Trustworthiness on KG level	Approaches of data curation and insertion.	Trustworthiness of data
Trustworthiness on statement level	Statements are written via provenance vocabulary.	Consistency of data type
Using unknown and empty values	Unknown and empty values are permitted.	Format of data
Intrinsic dimension – Consistency		
Check of schema restrictions during insertion of new statements	Examining schema restrictions before inserting new statements is required.	Consistency of data type
Consistency of statements – class constraints	Examining class restrictions before inserting new statements is required.	Consistency of data type
Consistency of statements – relation constraints	Examining relation restrictions before inserting new statements is required.	Consistency of data type

Table 4-2: Knowledge graph (KG) quality and evaluation framework (revised based on Färber et al., 2018) (continued)

Metrics	Description	Aspects of quality
Contextual dimension – Relevancy		
Creating a ranking of statements	Statements can be expressed from temporal perspective.	Trustworthiness of data
Contextual dimension – Completeness		
Schema completeness	Classes and relations meet the gold standard.	Completeness of data
Column completeness	The instances of a class have the same relations.	Consistency of data type
Population completeness	The KG covers a basic population of general knowledge.	Completeness of data
Contextual dimension – Timeliness		
Timeliness frequency of the KG	The KG is updated continuously.	Trustworthiness of data
Specification of the validity period of statements	The validity period of statements is disclosed	Format of data
Specification of the modification date of statements	The modification date is specified.	Format of data

Table 4-2: Knowledge graph (KG) quality and evaluation framework (revised based on Färber et al., 2018) (continued)

Metrics	Description	Aspects of quality
Representational data quality – Ease of understanding		
Description of resources	Human-understandable resources/tags are available.	Format of data
Labels in multiple languages	Resources/tags are available in multiple languages.	Completeness of data
Understandable RDF serialisation	Alternative RDF serialisation is available.	Format of data
Self-describing URIs	Descriptive URIs are available.	Format of data
Representational data quality – Interoperability		
Avoiding blank nodes and RDF reification	RDFs in the KG do not contain empty nodes.	Completeness of data
Provisioning of several serialisation formats	Alternative RDF format is available.	Format of data
Using external vocabulary	Vocabulary used to represent RDFs is consistent.	Consistency of data type
Interoperability of proprietary vocabulary	Vocabulary used to represent RDFs contains interlinks.	Format of data

Table 4-2: Knowledge graph (KG) quality and evaluation framework (revised based on Färber et al., 2018) (continued)

Metrics	Description	Aspects of quality
Accessibility dimension – Accessibility		
Dereferencing possibility of resources	RDFs can be resolvable via HTTP.	Consistency of data type
Availability of the KG	Requests to the KG are restricted.	Usefulness
Provisioning of public SPARQL endpoint	Complex queries are allowed via SPARQL.	Usefulness
Provisioning of an RDF export	RDF dump is available and downloadable.	Usefulness
Support of content negotiation	Ambiguous search is allowed.	Usefulness
Linking HTML sites to RDF serialisations	RDFs are represented in HTML links.	Format of data
Provisioning of KG metadata	Metadata of the KG is machine-readable.	Format of data
Accessibility dimension – License		
Provisioning machine-readable licensing information	The license information is machine-readable and available.	Format of data
Accessibility dimension – Interlinking		
Interlinking via owl:sameAs	Cross-KGs links are available.	Format of data
Validity of external URIs	External links are resolvable.	Format of data

Table 4-3: The result of evaluating knowledge graph (KG) candidates (based on the result of Färber et al., 2018)

Metrics	**DBpedia**	**Freebase**	**OpenCyc**	**Wikidata**	**YAGO**	**Consideration**
Intrinsic category – Accuracy						
Syntactic validity of RDF documents	1	1	1	1	1	High
Syntactic validity of literals	0.994	1	1	1	0.624	Low
Semantic validity of triples	0.990	0.995	1	0.993	0.993	High
Intrinsic category – Trustworthiness						
Trustworthiness on KG level	0.5	0.5	1	0.75	0.25	High
Trustworthiness on statement level	0.5	1	0	1	1	High
Using unknown and empty values	0	1	0	1	0	Medium
Intrinsic category – Consistency						
Check of schema restrictions during insertion of new statements	0	1	0	1	0	High
Consistency of statements – class constraints	0.875	1	0.999	1	0.333	High
Consistency of statements – relation constraints	0.992	0.451	1	0.500	0.992	High

Table 4-3: The result of evaluating knowledge graph (KG) candidates (based on the result of Färber et al., 2018) (continued)

Metrics	**DBpedia**	**Freebase**	**OpenCyc**	**Wikidata**	**YAGO**	**Consideration**
Contextual category – Relevancy						
Creating a ranking of statements	0	1	0	1	0	High
Contextual category – Completeness						
Schema completeness	0.905	0.762	0.921	1	0.952	Medium
Column completeness	0.402	0.425	0	0.285	0.332	High
Population completeness	0.93	0.94	0.48	0.99	0.89	Low
Contextual category – Timeliness						
Timeliness frequency of the KG	0.5	0	0.25	1	0.25	Medium
Specification of the validity period of statements	0	1	0	1	1	Low
Specification of the modification date of statements	0	1	0	0	0	Low
Representational data quality – Ease of understanding						
Description of resources	0.704	0.972	1	0.999	1	Low
Labels in multiple languages	1	1	0	1	1	Low
Understandable RDF serialisation	1	1	0	1	1	Low
Self-describing URIs	1	0.5	1	0	1	Low

Table 4-3: The result of evaluating knowledge graph (KG) candidates (based on the result of Färber et al., 2018) (continued)

Metrics	DBpedia	Freebase	OpenCyc	Wikidata	YAGO	Consideration
Representational data quality – Interoperability						
Avoiding blank nodes and RDF reification	0.5	0.5	0.5	0	0.5	Medium
Provisioning of several serialisation formats	1	0	0.5	1	1	Low
Using external vocabulary	0.61	0.108	0.415	0.682	0.134	High
Interoperability of proprietary vocabulary	0.150	0	0.513	0.001	0	Medium
Accessibility category – Accessibility						
Dereferencing possibility of resources	1	0.437	1	0.414	1	High
Availability of the KG	0.996	0.999	1	0.999	0.731	High
Provisioning of public SPARQL endpoint	1	0	0	1	1	Low
Provisioning of an RDF export	1	1	1	1	1	Low
Support of content negotiation	0.5	0	0	1	1	Medium
Linking HTML sites to RDF serialisations	1	1	0	1	1	Low
Provisioning of KG metadata	1	0	1	0	0	Low
Accessibility category – License						
Provisioning machine-readable licensing information	1	0	0	1	0	Low

Table 4-3: The result of evaluating knowledge graph (KG) candidates (based on the result of Färber et al., 2018) (continued)

Metrics	**DBpedia**	**Freebase**	**OpenCyc**	**Wikidata**	**YAGO**	**Consideration**
Accessibility category – Interlinking						
Interlinking via owl:sameAs	0.251	0	0.382	0	0.310	High
Validity of external URIs	0.929	0.908	0.894	0.957	0.956	High
Unweighted Average	0.683	0.603	0.496	<u>0.752</u>	0.625	
Weighted Average*	**0.675**	**0.622**	**0.560**	**<u>0.734</u>**	**0.616**	

*Assuming the weights of consideration are: High = 3, Medium = 2, Low = 1

4.1.2.4 Wikidata

Wikidata (formerly *Freebase*) is an open-sourced and collaboratively edited set of KG covering general knowledge. The expression of knowledge in Wikidata is based on ontological logic, including concept, property and value. The information provided by Wikidata can be interpreted by both humans and machines, allowing end users to conduct further analysis, such as enquiring via the Wikidata Query Service and Dumping Service to extend the Wikidata functions. Wikidata has drawn considerable attention in academia recently, especially in the biomedical and medical context. Many studies have considered Wikidata as a validated corpus for analysis and contributor to the knowledge map (Tharani, 2021; Turki et al., 2019; Waagmeester et al., 2020, 2021). Waagmeester et al. (2020) also validated that Wikidata, as continuously updated and community-maintained KG, meets the principles of Findability, Accessibility, Interoperability and Reusability (FAIR) and has strong potential to be applied to scientific research.

In the transportation context, some studies have applied publicly available KGs, such as DBpedia[4], as the foundation for their research purpose. Metzke et al. (2013) developed a semantic complex event processing for logistics to detect meaningful events (such as a flood) on a transportation route. DBpedia is used as a top-level ontology with slight modification. Abdullah et al. (2017) used DBpedia to retrieve the semantic meaning from the recorded voice of control tower operators and pilots to identify aviation safety-related events.

For the entity linking model, this thesis only leverages the knowledge network and Wikidata Query Service API to extract connections between entities of interest. The content,

[4] DBpedia is a project aiming to extract structured data from Wikipedia and publish them in RDF formats to provide a series of services around the extracted data, for instance, the SPARQL enquiring system and a number of mappings to external ontologies, an ontology itself.

explanation and other information of entities on Wikidata are not considered to simplify the process and ensure the consistency of data.

4.1.2.5 Tagme – the open-sourced on-the-fly annotation of short text

To extract potential entities in text, the off-the-shelf API *Tagme* is introduced to address the MD and CG question. *Tagme* is a toolkit for augmenting structure-free text by connecting text to the relative knowledge hyperlinks under the Wikipedia framework, enabling users to recognise the sequence of terms and assign an unambiguous entity to identified terms without training a supervised learning NLP model. Wikipedia pages are selected for the catalogue used to categorise the entities due to the high-quality entities and strong networked structure (Ferragina & Scaiella, 2010; Hasibi et al., 2016). The *Tagme* toolkit also produces flexible outcomes allowing end users to conduct further analysis.

For the MD, the *Tagme* compares the input text with all Wikipedia pages exclusive of disambiguation pages, redirect pages and other irrelevant pages, and extracts a set of potential entities A_t for each term that might be connected to the Wikipedia entity. Subsequently, the probability of generated candidates is calculated based on *relativeness*, a metric measuring the overlap between the in-linking pages in Wikipedia. The voting scheme is used in Tagme, a concept of pairwise comparison between candidates. The voting process can be illustrated as Equation 4-9.

$$vote_b(p_a) = \frac{\sum_{p_b \in Pg(b)} rel(p_b, p_a) \times Pr(p_b|b)}{|Pg(b)|} \qquad \text{Equation 4-9}$$

where *a, b* denotes two Wikipedia entities with the page p_a and p_b and ambiguous set of pages $Pg(a)$ and $Pg(b)$. The $rel(p_b, p_a)$ represents the *relativeness* of p_a and p_b. The $vote_b(p_a)$ refers to the possibility that instead of assigning to the entity *a*, the entity *b* has more chances of being correct. Note that if only one candidate is identified during the CG, the $Pr(p_b|b)$ and $|Pg(b)|$ are equal to 1 so the $vote_b(p_a)$ is equal to $rel(p_b, p_a)$, implying the entity does not have more than one meaning.

However, the outcome of *Tagme* might still be ambiguous, and the CG process does not consider the context-wide feature. For instance, considering the short sentence, “The near miss was reported at 15:32”, *Tagme* identifies the near miss as a popular rock band rather than an incident. To address this issue, this thesis further extends the CG process by proposing the context-sensitive disambiguation process.

4.1.2.6 The context-sensitive disambiguation process

Consider a collection of entities identified by *Tagme* from a set of documents, and each entity has been linked to a Wikipedia page. For each Wikipedia page, several hyperlinks are connected to the mentioned in-text entities linking to other Wikipedia pages. For instance, the text in the Wikipedia page “Classification of railway accidents” must contain a hyperlink “rail” directing to another Wikipedia page “Rail transport”. We extract all hyperlinks contained in all linked Wikipedia pages and construct a network with dots representing Wikipedia pages and edges describing hyperlinks between pages.

Next, the Degree of Centrality (DoC) and Eigenvector Centrality (EC) based on the graph theory is introduced to calculate the importance of each dot (Bonacich, 2007; Kwayu et al., 2021; Lin et al., 2021). Assuming that railway accident reports have substantial homogeneity in the same context, the most relevant entity is expected to be linked many times on the network. For example, the entity “near miss” for the incident is expected to be linked with the hyperlink more than the entity “Near Miss” for the rock band.

The DoC is a conventional measure counting the number of edges one dot has (Hansen et al., 2019). For standardisation, the DoC of one dot v is calculated as Equation 4-10.

$$DoC_v = \sum_{t=1,t\neq v}^{M(v)} \frac{P_{min,v,t}}{M(v)-1} \qquad \text{Equation 4-10}$$

where $M(v)$ is the set of neighbours of the dot v, $\mathrm{P}_{\mathrm{min,v,t}}$ represents the minimum number of transfers between dots *v* and *t*.

On the other hand, the EC of one dot v is calculated based on the centrality of its neighbours (Equation 4-11).

$$EC_v = \frac{1}{\lambda} \times \sum_{t \in M(v)} EC_t \qquad \text{Equation 4-11}$$

where $M(v)$ is the set of neighbours of the dot v and λ is a constant representing the eigenvalues of the graph.

The entity on the network with the highest DoC is selected as the solution for disambiguation. If more than one entity has the highest DoC, then the entity with the highest EC is selected. However, if more than one entity has the same DoC and EC or the DoC and EC of all entities are 0, then these entities are reviewed manually before selecting the solution of disambiguation. An overview of the process is in Figure 4-10.

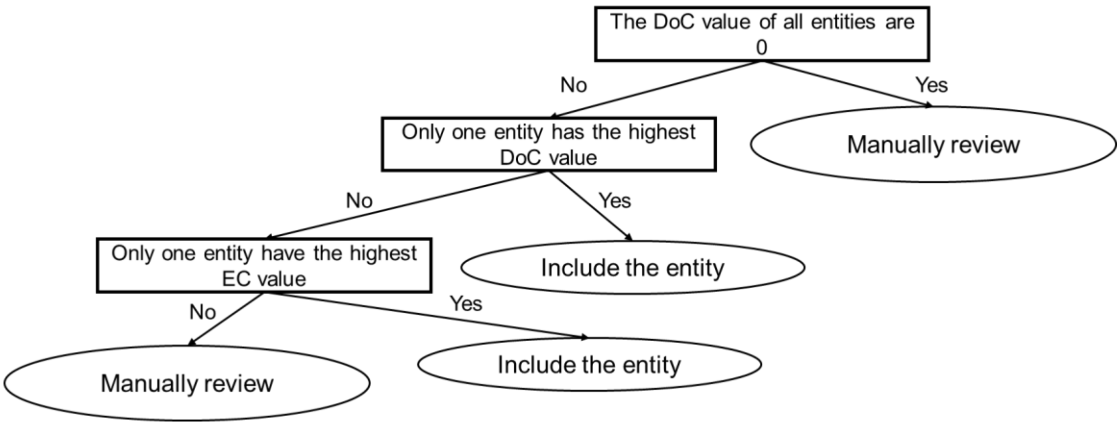


Figure 4-10: Illustration of context-sensitive disambiguation process

4.1.3 Covariate analysis

The STM advances topic modelling by introducing the covariate and sensitivity analysis, allowing users to understand the effect of a factor on topics. There are three types of covariate analysis offered by the STM: *Difference*, *Point estimate* and *Continuous*.

Difference analysis estimates the mean difference in topic proportions among two different values of the binary covariate, revealing where the selected word is prevalent among all topics. Taking the entity “Network Rail” as an example, we first search the occurrence of the entity as a covariate and identify topics containing high proportions of this entity compared with other text with statistical significance. All the selected topics are considered to have high relation with the entity “Network Rail” (Figure 4-11).

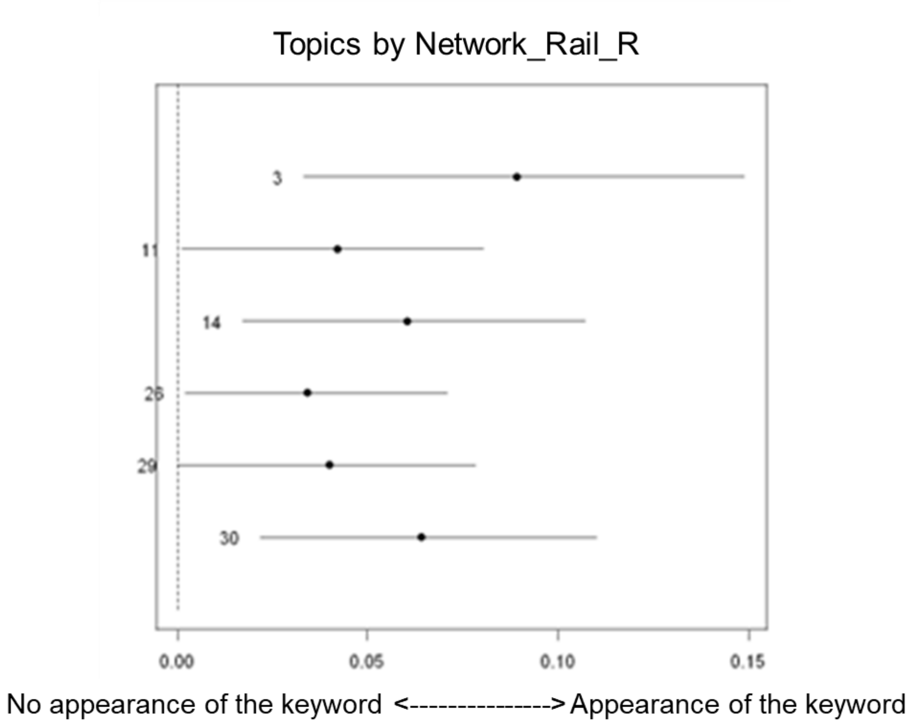


Figure 4-11: A demonstration of difference covariate analysis using Network Rail as the factor

Point estimate analysis measures the mean topic proportions for each value of the covariate, revealing the different proportions of each categorical covariate on a specific topic. An example is illustrated as Figure 4-12. Suppose we create a factor with three categories: high, medium and low appearance of the keyword “Network Rail”. Then the Point estimate covariate analysis can show that most of the documents sorted to topic 3 mention the keyword at a higher level of occurrence.

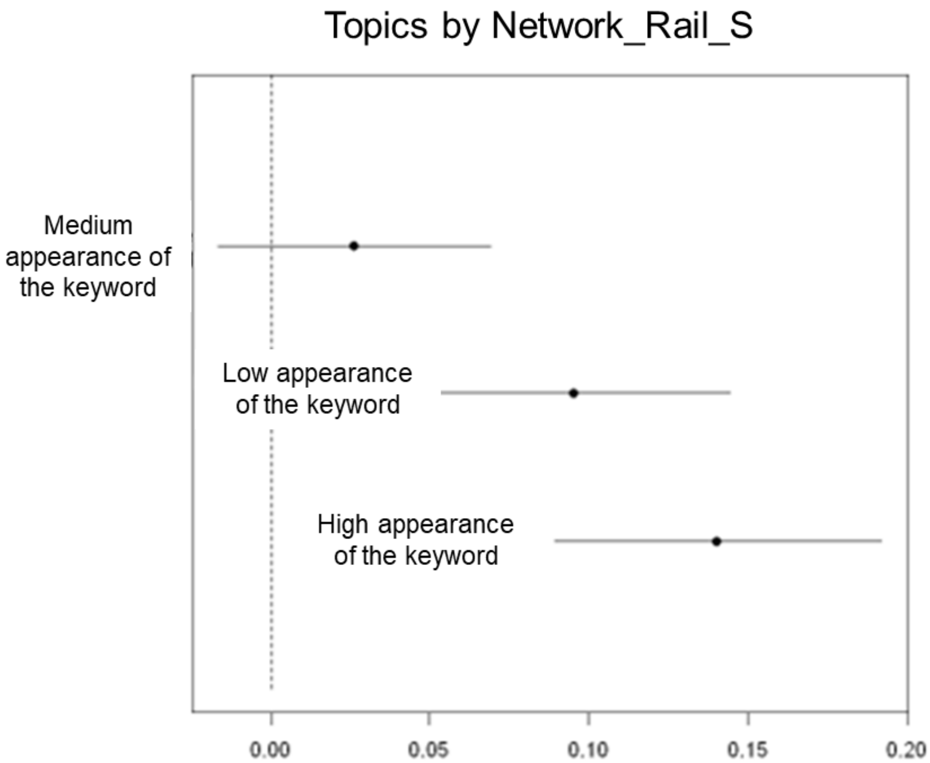


Figure 4-12: A demonstration of point estimate covariate analysis using Network Rail as the factor

Continuous analysis estimates how topic proportions influence over a continuous covariate, revealing the trend of the effect brought by selected keywords on the topic. Figure 4-13 shows an example of *Continuous* covariate analysis over topics 3 and 30. The occurrence of keyword “Network Rail” has a positive correlation with the documents assigned to topic 30 but has a negative correlation with topic 3. This might imply the keyword “Network Rail” is frequently used on documents assigned to topics 3 and 30, but a considerably higher frequency is found in only topic 30.

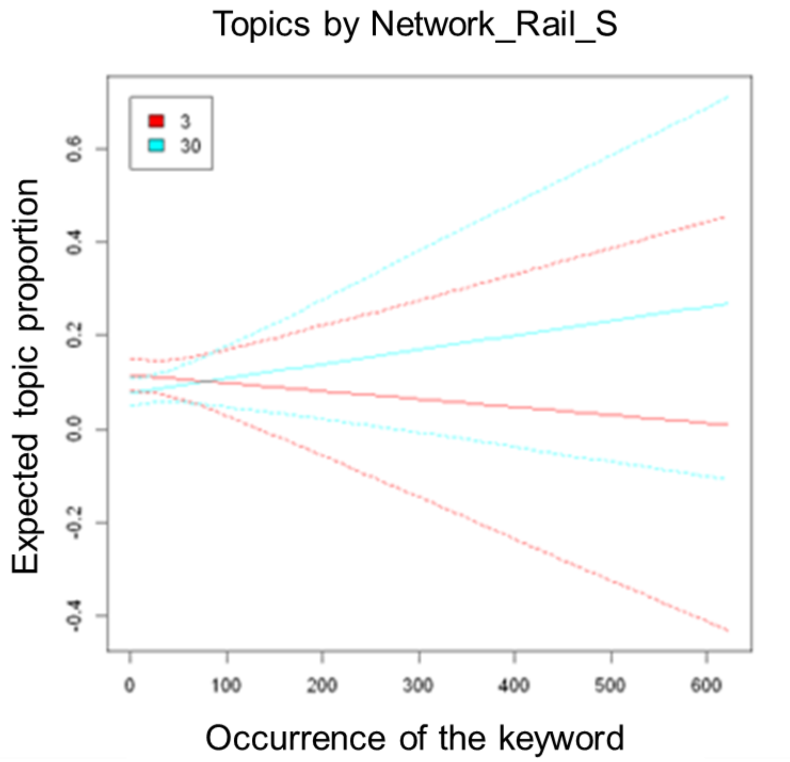


Figure 4-13: A demonstration of continuous covariate analysis using Network Rail as the factor

4.1.4 Temporal analysis

The STM can also reveal the prevalence of topics over time. Figure 4-14 depicts the trend of the recommendations related to “sharing knowledge with organisations” made by ATSB over time (see Section 5.2 for more details of the ATSB dataset). The solid line represents the mean, and the dotted lines are the standard deviation.

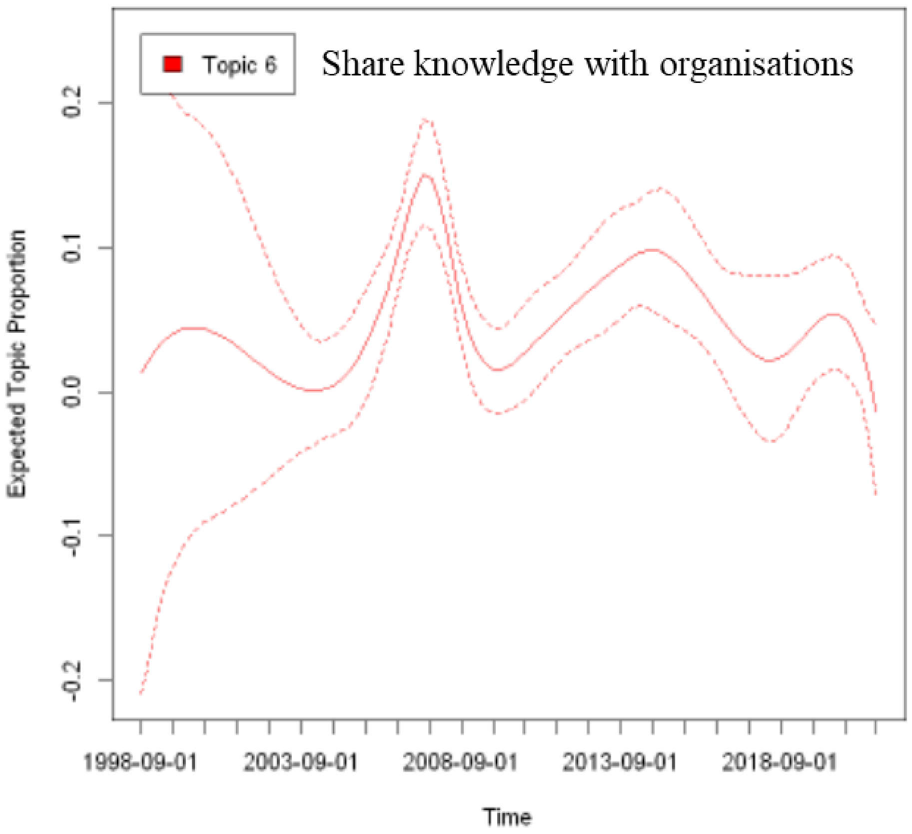


Figure 4-14: A demonstration of temporal analysis using the recommendation dataset of ATSB reports

4.2 Evaluation of the model – the scoping workshop and survey

Despite the mathematical indicators proposed in this chapter for measuring the performance of the developed model, the evaluation of the usefulness and consistency of the railway practice still required the involvement of experts. Two online workshop sessions were arranged, each of them comprised two main topics: organisational interaction within the railway industry in the context of receiving, processing and disseminating rail safety-related knowledge; and feedback on this research and the potential for further implementation. Specifically, we aim to achieve the following objectives:

Part I:

(1.) Confirming how railway accident reports and investigations influence the industry's day-to-day operation.

(2.) Understanding how rail safety-related knowledge is obtained, processed and disseminated by the industry.

(3.) Understanding how the railway industry improves its safety system in the light of railway accident reports, and who or what would be involved. Additionally, would be the lessons learned by the industry?

(4.) Understanding what type of information the railway industry needs when improving railway safety and how each participant's role affects the needs.

Part II:

(1.) Presenting the current work and results and collecting feedback.

(2.) Understanding how the participants interpret the results from the model and discovering the pros and cons from the industry's perspective.

The workshops were held virtually. All content, including the discussion and the survey before, during and after this workshop, was recorded but anonymised. Since the initial number of expressions of interest was below 10 during private invitation, so open registration was in place and the threshold of minimal work experience was lowered to 1 year. All participants involved in four countries (the UK, Australia, China and Taiwan) were unidentifiable in both the recording and further analysis. The details of the workshops are shown in Table 4-4. The process of invitation started with asking for expressions of interest (EOI) from specialist organisations The workshop and the survey were approved by the Human Research Ethics Committee at the University of Sydney (the approval letter can be found in Section 9.2). All responses were entirely confidential. Data processing and secure storage followed the

University's research code of conduct and research data management policy. All comments and suggestions were welcome before, during and after the workshop through email. For the convenience of participants from different time zones, there were two identical sessions. The content of each session was the same. One session was chaired by Dr Geoffrey Clifton (Senior Lecturer, ITLS) and one by Professor John Nelson (Chair in Public Transport, ITLS). The pilot survey was distributed to and reviewed by two academics for evaluation. Minor modifications relating to the structure of the questions and terminology used were amended.

Table 4-4: Details of the two workshops

	Session 1	**Session 2**
UTC-time	Friday, 8 April 2022, 10:00:00	Monday, 11 April 2022, 23:00:00
New York	Friday, 8 April 2022, 06:00 – 07:30	Monday, 11 April 2022, 19:00 – 20:30
London	Friday, 8 April 2022, 11:00 – 12:30	Tuesday, 12 April 2022, 00:00 – 01:30
Sydney	Friday, 8 April 2022, 20:00 – 21:30	Tuesday, 12 April 2022, 09:00 – 10:30
Wellington	Friday, 8 April 2022, 22:00 – 23:30	Tuesday, 12 April 2022, 11:00 – 12:30

The information was expected to be extracted through the roundtable discussion, Q&A session after the presentation and a take-out survey. Two types of exercises were included to extract the participants' judgment. The first exercise was to ask participants to determine the topic's content by providing the keywords or the extracted articles from the topic model. The second exercise was to ask participants to judge the pre-set name of a group of articles. These exercises were the main session to evaluate the performance of the developed model from the perspective of railway practice. The lead researcher presented how the model was developed and what the outcome looked like, followed by a roundtable discussion relating to how participants interpreted the result of the model, the consistency of the presented knowledge and further potential application in practice. Outcomes of the discussion is summarised in Section 5.6.1.

4.3 Synthesis of findings

This chapter has synthesised the development of models for the automation of railway accident analysis. The proposed analysis processes, including the topic modelling, entity linking strategy, covariate analysis, temporal analysis, and the evaluation of developed models, have been explained. Potential candidates and the selected approach for use in this thesis were compared.

For the topic modelling, the mechanism and the used NLP model of each approach were illustrated. A comparison between candidates from the perspective of dimensionality of processed data was provided and analysis toolkits and outcomes were described. BERTopic was selected to analyse railway accident reports at the sentence level due to the ability to capture higher semantic features and the compatibility of analysing short to medium-length of text, whereas the STM was selected to analyse railway accident recommendations at the sentence level given the relatively high level of homogeneity on the writing style of recommendations and limited semantic features. Demonstrations on applying the covariate analysis and temporal analysis were also provided.

The entity linking strategy in the study involved the creation of a railway accident ontology by collecting existing ontologies and reviewing accident reports. Knowledge graphs were used to identify and disambiguate entities from these reports. An API toolkit *Tagme* is used to identify additional entities, and a context-sensitive disambiguation process is proposed to enhance entity detection and linking. The main purpose of the ontology is to standardise terminology and connect identified keywords from topic modelling with original accident reports, making the keywords more interpretable by linking them to relevant entities.

Last, a scoping workshop and survey were described. The workshops were arranged with industry experts to evaluate the developed model. The roundtable discussion, Q&A session after the presentation and a take-out survey were used to collect criticisms and comments from the perspective of the railway industry. In addition, participants' interpretations, judgments and expectations were extracted and observed.

5. Initial analysis and topic modelling

This chapter illustrates the application of the methods presented in Chapter 4 and develops novel frameworks for depicting the nature of hazards, railway accidents and recommendations from railway accident reports with real-world case studies. The structure of the chapter is as follows: the process of data acquisition is discussed (Section 5.1), followed by the data pre-processing (Section 5.2). The overview of individual country analysis is presented (Section 5.3), and the ontology, knowledge graph selection and entity linking are elaborated (Section 5.4). Next, the cross-country analysis is discussed (Section 5.5), followed by the evaluation of model through workshops and survey (Section 5.6). Last, the synthesis of findings is presented.

5.1 Data acquisition

To demonstrate the application of the proposed models, railway accident reports published by independent railway accident investigation bodies from four countries are used: Australia, Canada, the UK and the US. Railway accident reports complied by independent railway accident investigation organisations are regulated by a national legal framework and provide unbiased and blame-free details for promoting a railway safety culture. Despite the differences in writing styles and terminology used, all reports consist of the summary of the accident, the analysis, the investigation, key findings, conclusions and recommendations (if applicable). The database provided by investigators covers various periods of time. For the best understanding of railway accident knowledge and model performance, all retrievable railway accident reports in PDF format from the official websites of the investigated countries were retrieved. Data from the ATSB and TSB was retrieved from their websites directly because the full text is provided and crawlable via HTML. Scanned files were removed due to the technical difficulties of recognising the text. The following steps were applied to clean the data:

(1.) Remove the pages of the preface, cover, table of contents, appendices, and other irrelevant pages.

(2.) Remove sentences only for structure purposes, such as headers, footers, page numbers, annotations, and other irrelevant content.

(3.) Remove tables and figures.

(4.) Split the data into sentences.

(5.) Remove sentences with no analysable information. Sentences with at least one of the following features were removed (Fischbach et al., 2020):

   a. The sentence starts with or contains “Figure” or “Table”.
   b. The sentence starts with a page character or “Chapter”.
   c. The sentence consists of less than 50 characters and does not end with “.”, “?” or “!”.
   d. The sentence consists of a group of at least 4 successive “.” characters (deletion of entries in the table of contents).

(6.) Remove sentences which begin with “Note” but are not followed by “that” (Fischbach et al., 2020).

(7.) Remove text that begins with “(figure …” and ends with “..)” in sentences.

(8.) Remove the fronter and footer.

(9.) Extract recommendations as another independent dataset[5].

Original reports were split into the railway accident dataset and the recommendations dataset at the sentence level. Table 5-1 shows the overview of the processed railway accident dataset. The RAIB published a series of review reports, such as the *Investigation into the safety of automatic open level crossings on Network Rail’s managed infrastructure* (RAIB, 2011). These reviews overlapped with published reports and were excluded from the dataset. The

[5] The NTSB and RAIB provide an independent recommendations dataset in an editable form and it was used directly.

ATSB and TSB provide the full text of railway accident reports on their websites and reports were retrieved directly. Despite the availability of early reports published by the NTSB, scanned files were retrieved and they are excluded from the dataset due to the difficulties of converting them into editable text.

Table 5-1: Overview of the processed railway accident dataset

Body	No. of reports	No. of sentences	Period covered	Note
RAIB	339	124,990	2005–2019	Review reports are removed.
ATSB	250	84,679	1999–2021	Reports are retrieved from website directly.
NTSB	274	92,406	1996–2021	Reports earlier than 1996 are scanned files.
TSB	415	104,720	1993–2021	Reports are retrieved from website directly.
Total	1,278	406,795		

Table 5-2 summarises the processed recommendations dataset at the sentence level. The number of recommendations made by the TSB is limited because the TSB only publishes recommendations at the highest level of accidents with severe consequences. Similar circumstances can be found in the ATSB dataset given that only identified risks are highlighted without publishing recommendations directly, offering the railway industry the flexibility to propose strategies for managing risk factors. Another note is that the NTSB provides an independent recommendations dataset ranging from 1966 to 2020 and stored in an editable way. Therefore, all recommendations are retrieved to understand the composition of recommendations across time.

Table 5-2: Overview of the processed recommendations dataset

Body	No. of sentences	Period covered	Note
RAIB	4,807	2005–2019	All reports are linked to corresponding recommendations.
ATSB	1,074	1999–2021	Only a limited number of reports lead to recommendations.
NTSB	3,185	1966–2020	Reports earlier than 1996 are scanned files, but the recommendations dataset is independent, editable and retrievable from 1966 to 2020.
TSB	76	1991–2021	Only a limited number of reports lead to recommendations.

5.2 Data pre-processing

The data pre-processing is only applicable to the STM because the performance of the clustering model based on bag-of-word embedding methods strongly relies on the data quality. The noise, such as the duplication of content or flaws in raw data, would be clustered by the STM as well during the analysis and influences the outcome significantly. On the other hand, BERTopic uses the Sentence-BERT (SBERT) framework, allowing users to convert the input data into dense vectors with pre-set processes and pre-trained language models without manual intervention (Reimers & Gurevych, 2019). The noise is kept in the data at this stage but will be identified and labelled as outliers during follow-up clustering approaches.

The data pre-processing for the STM consists of the following steps: lowercasing, digital number removing, punctuation removing, and stemming. The R package *textProcessor* is implemented in the STM. The metadata is associated with processed text by the *quanteda* package, converting data into a document-term matrix and holding covariates at the document level (Benoit et al., 2018). Other libraries under *quanteda* also provide a wide range of functions, such as reading data in multiple forms. The output can be directly fit into STM functions.

5.3 Overview of individual country analysis

The following section provides an overview of analysis results from the four investigated countries using the BERTopic model for topics and the STM for recommendations.

5.3.1 RAIB, UK – BERTopic model for topics

To establish an appropriate BERTopic model for the RAIB dataset, the number of topics needs to be determined by selecting a proper minimum cluster size (MCS). The larger cluster size leads to a small number of topics and *vice versa*. There is no common agreement on determining the best MCS yet in the literature. Thus, this study uses the coherence value and the distribution of the number of sentences over each topic to determine the MCS. Figure 5-1 and Figure 5-2 show the coherence score over different MCS values and the distribution of the number of sentences over each topic with varying values of MCS on the RAIB dataset, respectively. The first topic in Figure 5-2 refers to the outlier group. The more sentences sorted to the first group indicate that more outliers are identified under the MCS value. The coherence score drops significantly after the MCS reaches 25.

On the other hand, the number of outliers also increases dramatically as the MCS value rises. Although MCS 5 has a lower number of outliers and higher coherence score, an excessive number of topics and higher homogeneity might occur and disrupt the following topics' interpretation given the low threshold of making a group is set. Instead, MCS 15 has a relatively high coherence score and the second-lowest number of outliers. Furthermore, the distribution of the number of sentences over each topic is similar to other higher MCS values, implying a balanced topic distribution. Therefore, the MCS is set to 15 to build the BERTopic model for the RAIB dataset.

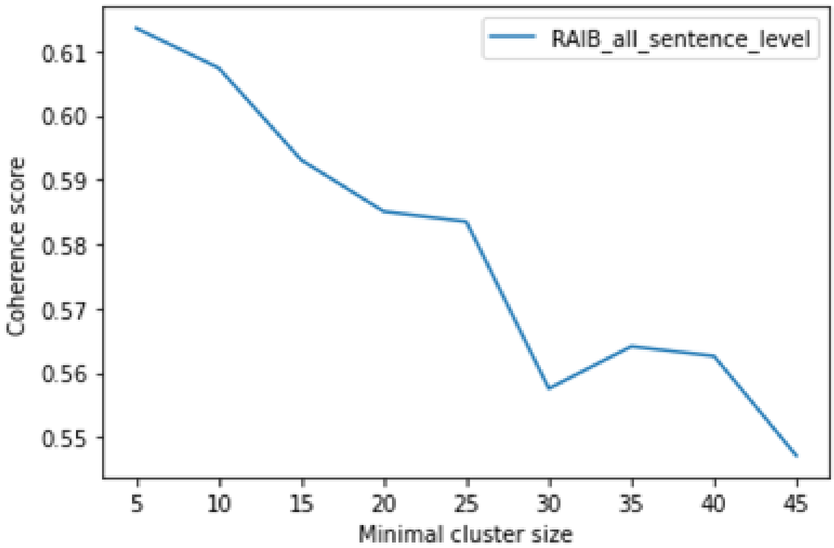


Figure 5-1: Coherence score over different minimal cluster size (MCS) values (RAIB)

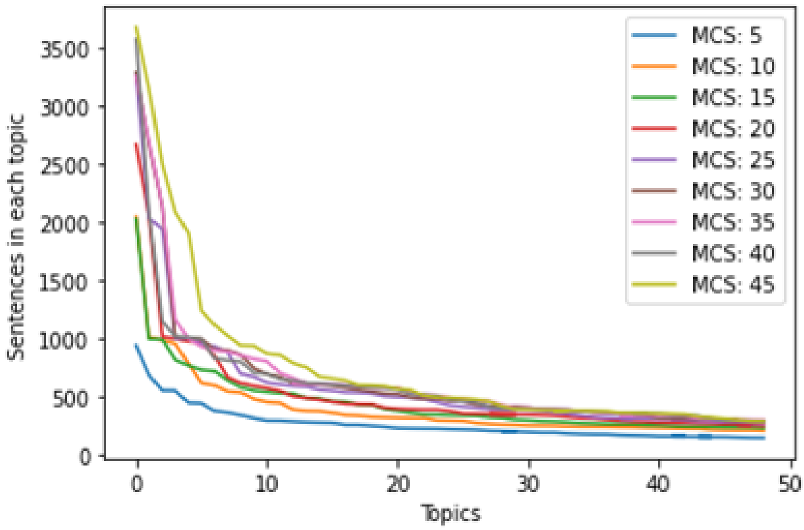


Figure 5-2: Distribution of the number of sentences over each topic with different minimum cluster size (MCS) values (RAIB)

Table 5-3 shows the extracted topics with high occurrence and associated keywords after removing irrelevant topics. The name of each local topic is assigned in accordance with the identified keywords and their occurrence probabilities. For instance, the top 5 keywords with the highest occurrence probability in topic 6 are “mph”, “kmh”, “speed”, “maximum” and “permitted”, indicating that sentences clustered into this topic are relevant to the topic “speed”. Therefore, the topic name “speed” is assigned to topic 6. In addition, some topics at a lower level (local topics) have correlation or causality and might be integrated to describe a concept at a higher level (interval and global topics). For example, topic 6 and topic 42 seem to have causal relation because the emergency brake might be triggered by speeding. Therefore, the name of the interval topic for topics 6 and 42 is given as "emergency brake at high speed". To aggregate multiple interval topics and generate the interpretation of events during the railway accident, several relevant interval topics are further integrated to describe the causal chain of events. The BERTopic model can illustrate the distance between topics via the inter-topic distance diagram (Figure 5-3) and provide a potential combination of latent topics for understanding the nature of railway accident types. For instance, topics 6, 42, 27, 11 and 30 are integrated to describe trains that apply the emergency brake at high speed due to the Automatic Warning System (AWC) isolation or work site hazards.

However, extracted topics might be ambiguous or irrelevant to the railway accident because the input data is set at the sentence level. Thus, the coefficient of variance (CV) is introduced to address this issue and understand the distribution of topics over documents. The CV is calculated as the following:

$$CV = \sigma_i^2 / \mu_i \qquad \text{Equation 5-1}$$

where σ_i^2 refers to the variance of sentences under topic *i* over documents and μ_i refers to the mean of the number of sentences under topic *i* over documents. The higher CV indicates the topic has a uniform distribution over documents, implying a general topic that can be commonly found in railway accident reports. For instance, the topic 6 "speed" in Table 5-3 has a CV value of 2.85, showing that the speed of objects involved is a commonly investigated element in railway accidents. Another example is topic 27 in Table 5-3 with a CV value of 137.83, suggesting most sentences under this topic come from a small number of documents with similar characteristics.

In the RAIB dataset, several distinct topics are identified with high CV values at interval or global levels such as platform–tram interface (topic 0-1), track–wheel interface (topic 17), and work site safety and hazards with engineering units. Figure 5-3 also shows the distribution of identified intervals and global topics. The x-axis and y-axis are the dimensions that the BERTopic model extracts for clustering and only used for computer calculation purposes.

Table 5-3: Topic descriptions and coefficient of variance (CV) of RAIB dataset

<table>
<tr><th>Topic</th><th>Topic – local</th><th>Topic – interval</th><th>Topic – global</th><th>CV</th></tr>
<tr><td>6</td><td>Speed</td><td rowspan="2">Emergency brake at high speed</td><td rowspan="6">Trains apply emergency brake at high speed due to 1) AWC isolation, 2) work site hazards</td><td>2.85</td></tr>
<tr><td>42</td><td>Emergency brake</td><td>2.63</td></tr>
<tr><td>27</td><td>Automatic Warning System (AWS) isolation/ active</td><td rowspan="2">AWS isolation due to error warning/ failure of signal system</td><td>137.83</td></tr>
<tr><td>11-1</td><td>Sounded horn</td><td rowspan="2">9.68</td></tr>
<tr><td>11-2</td><td>Train horn</td><td rowspan="2">Work site safety and hazards – site workers</td></tr>
<tr><td>30</td><td>Site lookout</td><td>18.20</td></tr>
<tr><td>12</td><td colspan="2">CCTV, monitor/recording</td><td rowspan="3">Unawareness of platform–tram interface or pedestrians on track/ level crossing due to fatigue or incomplete monitoring system</td><td>22.01</td></tr>
<tr><td>1</td><td colspan="2">Fatigue</td><td>71.97</td></tr>
<tr><td>0-1</td><td>Trams / pedestrian</td><td>Platform–tram interface/ striking pedestrian</td><td rowspan="2">164.65</td></tr>
<tr><td>0-2</td><td>Trams / Sandilands*</td><td colspan="2">Tram-specified accident (i.e., Overturning)</td></tr>
<tr><td>2</td><td>Communication – Signaller (radio, GMR-S, etc.…)</td><td colspan="2" rowspan="2">1. Signaller–driver interface
2. Staff training / knowledge</td><td>21.21</td></tr>
<tr><td>4</td><td>Driver knowledge, training, instruction</td><td>15.17</td></tr>
<tr><td>13</td><td>Time</td><td colspan="2" rowspan="2">Background information</td><td>1.90</td></tr>
<tr><td>25</td><td>Location</td><td>1.66</td></tr>
</table>

Table 5-3: Topic descriptions and coefficient of variance (CV) of RAIB dataset (continued)

<table>
<tr><th>Topic</th><th>Topic – local</th><th>Topic – interval</th><th>Topic – global</th><th>CV</th></tr>
<tr><td>17</td><td>Sanding / adhesion</td><td>Track–wheel interface</td><td rowspan="2">Relation between set of units and track–wheel interface</td><td>110.49</td></tr>
<tr><td>20</td><td colspan="2">Set/type of train (single, multiple, diesel, electric unit)</td><td>4.41</td></tr>
<tr><td>15</td><td>Suspension system (bogie, wheel…)</td><td colspan="2" rowspan="2">Cause and result of flange climbing</td><td>10.91</td></tr>
<tr><td>36</td><td>Contact between flange and gauge</td><td>5.39</td></tr>
<tr><td>28</td><td>Deaths and injuries</td><td colspan="2" rowspan="2">Consequence of accidents</td><td>4.44</td></tr>
<tr><td>40</td><td>Property loss</td><td>0.70</td></tr>
<tr><td>48</td><td colspan="2">Grinding repairs</td><td rowspan="3">Track inspection/ recording/ maintenance</td><td>89.55</td></tr>
<tr><td>14</td><td>Track maintenance/inspection</td><td rowspan="2">Track defects inspection</td><td>15.05</td></tr>
<tr><td>49</td><td>Track geometry faults</td><td>13.05</td></tr>
<tr><td>39</td><td>Bolts failure</td><td rowspan="2">Design failure of the switch</td><td rowspan="8">1. Failure of signalling system
2. Failure of on-board equipment
3. Failure of infrastructure</td><td>96.27</td></tr>
<tr><td>47</td><td>(Nonadjustable) Stretcher bar</td><td>216.13</td></tr>
<tr><td>32</td><td>Wire–pantograph interface</td><td rowspan="2">Faults of wire–pantograph interface and inactive power system protection</td><td>111.29</td></tr>
<tr><td>37</td><td>Power system protection (circuit breaker)</td><td>47.98</td></tr>
<tr><td>29</td><td>Failure mode of the axle</td><td>Axle</td><td>315.86</td></tr>
<tr><td>33</td><td>Holdfast panel–sleeper interface</td><td>Level-crossing infrastructure</td><td>296.53</td></tr>
<tr><td>18</td><td>Switch interlocking system</td><td>Signalling system</td><td>169.99</td></tr>
<tr><td>46</td><td>Obstacle detection of doors</td><td>Door system</td><td>14.15</td></tr>
</table>

Table 5-3: Topic descriptions and coefficient of variance (CV) of RAIB dataset (continued)

Topic	Topic – local	Topic – interval	Topic – global	CV
9	Weather conditions	Natural disasters		4.22
10	Natural hazards (landslip, flood…)			35.86
22	PICOP (Person in Charge of Possession)	Work site safety and hazards – on-track possession	Work site safety and hazards with engineering units	28.31
43	SSOW (Safe System of Work)			21.99
44	Engineering units – RGU (Rail Grinding Unit)	1. Conditions of engineering units 2. Incidents and recommendations relating to engineering units		237.67
3	Engineering units – RRV (Road Rail Vehicles)			110.58
21	Engineering units – track trolley			75.05
45	Drugs and alcohol test**	Drug and alcohol conditions of the staff		3.04
35	Shunters/shunting activity	Hazards and regulations relating to shunters		46.08
16	Fire hazards	Fire incidents and response		437.23
31	Emergency service systems			12.31
19	COSS (Controller of Site Safety), driver	COSS–driver interface	Work site safety – COSS–driver interface and planning	10.47
41	COSS and site safety planning	Work site safety and hazards – site workers		5.23

Table 5-3: Topic descriptions and coefficient of variance (CV) of RAIB dataset (continued)

<table>
<tr><th>Topic</th><th>Topic – local</th><th>Topic – interval</th><th>Topic – global</th><th>CV</th></tr>
<tr><td>7</td><td>Declarative – risk assessment</td><td colspan="2" rowspan="2">Recommendations made by the RAIB on hazards identification and risk assessment</td><td>2.32</td></tr>
<tr><td>34</td><td>Hazards identification and risk assessment</td><td>3.26</td></tr>
<tr><td>38</td><td>Earthworks***</td><td rowspan="2">Infrastructure maintenance strategy and further improvement</td><td rowspan="3">Incidents and recommendation relating to Network Rail</td><td>28.90</td></tr>
<tr><td>5</td><td>Recommendations for Network Rail</td><td>10.68</td></tr>
<tr><td>24</td><td colspan="2">Network Rail's safety issues</td><td>2.42</td></tr>
</table>

* The occurrence of such a specific keyword is partly because of high level of reference.

** Although this topic may imply the existence of drug and alcohol hazards, we can only infer that this is an issue of concern for the RAIB and will always conduct a test after the occurrence of incidents.

*** According to the RAIB, earthworks refers to "a collective term for cuttings, embankments and natural slopes".

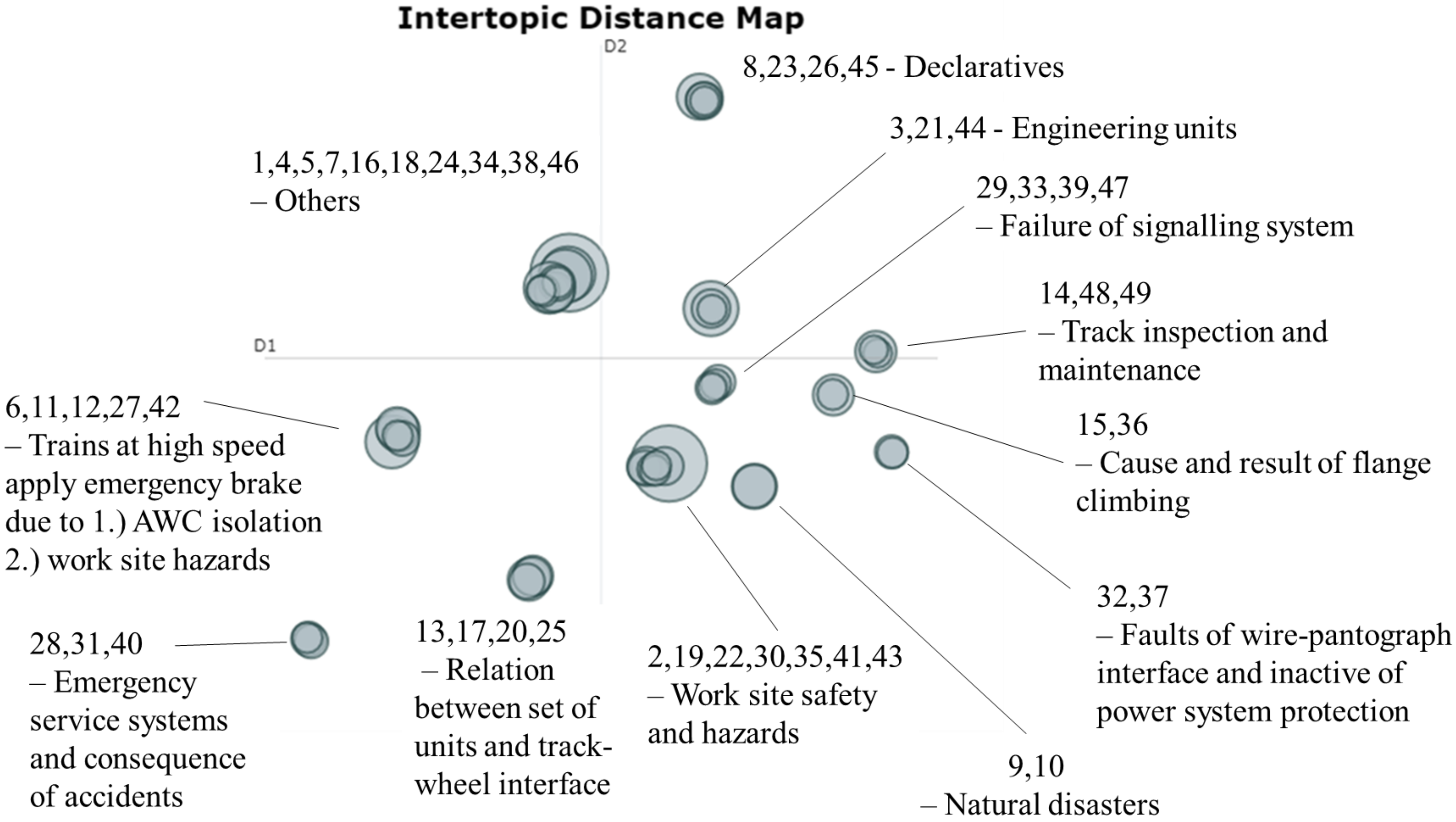


Figure 5-3: Inter-topic distance map of identified topics of the RAIB dataset

5.3.2 RAIB, UK – STM for recommendations data

For the STM model, the number of topics influences the performance of the model and needs to be estimated carefully. An iterative analysis sets the topic numbers from 5 to 50 and each model's performance is recorded using Equation 4-4. The results shown in Figure 5-4 suggest that using 26 topics for the RAIB recommendation dataset results in the best performance. Figure 5-5 illustrates the extracted keywords with the highest occurrence probability and the assigned name of each topic. The interpretation is completed by reviewing keywords and representative sentences from the perspective of recommendations for railway accidents rather than the nature of railway accidents. For instance, although the keywords "cross", "user" and "level" in topic 22 might refer to the mechanism of level crossing accidents, the name "review of consideration of design and standard for level crossing safety" is assigned to highlight the representation of other keywords and the real meaning of sentences sorted to this topic. Several identified topics of recommendations have been widely discussed in the railway accident studies and recommendation analysis in the literature, such as removal of the hazard (assessment and measurement), enhancement of design, enhancement of design assurance and approvals, steps to address safety culture (attitudes and behaviours), management process, enhancement of procedures, and training and competency (Braut et al., 2014; Cedergren & Petersen, 2010; Hulme et al., 2019; Tretten & Candell, 2021; Zhan & Zheng, 2016). However, other topics including standardisation of process and operation, cooperation, lesson learned processes, and documentation are seldom discussed.

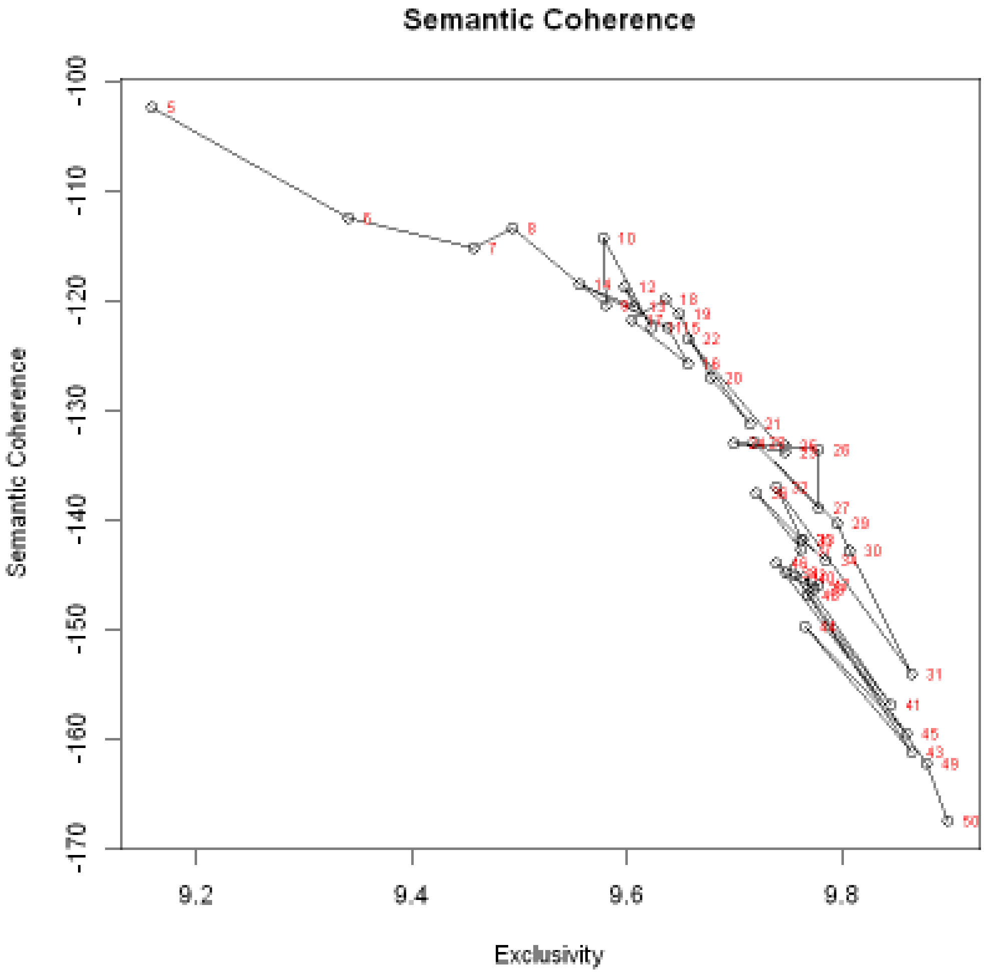


Figure 5-4: Sematic coherence and exclusivity score for each topic number (RAIB)

RAIB recommendation dataset

16. Report recommendations are implemented
5. Identify and implement appropriate measures for monitoring
22. Review of consideration of design and standard for level crossing safety
13. Include additional consideration in existing guideline/ assessment
2. Review training processes with relative organisations
7. Re-brief or re-train all staff involving in critical instructions
14. Report recommendations are satisfied
21. Review routine competence management and assessment (with other organisations)
20. Review consideration of consistent test for various conditions
9. Increase the awareness of risks in long possessions or work sites
12. Enhance existing processes for assessment
25. Update the process or guidance of change made
18. Review and improve physical equipment
24. Review and amend the processes and guidance applicable to Standards Committees
15. Modify the process for inspection and maintenance
3. Review communication protocols with relative organisations for lesson learnt
1. Review the design of rolling stocks toward different weather/ track conditions
4. Concerns of unaddressed issues
11. Guide on/ highlight particular hazards
19. Review associated rules and training documentations (with specialists)
6. Reconsideration of unnecessary decisions
23. Status of implementation
8. Review new processes with other relevant sources
17. Review the operation of the Overhead Line Electrification (OLE) system
10. Review principles of travelling near the maximum permitted line speed
26. N/A (Residentials)

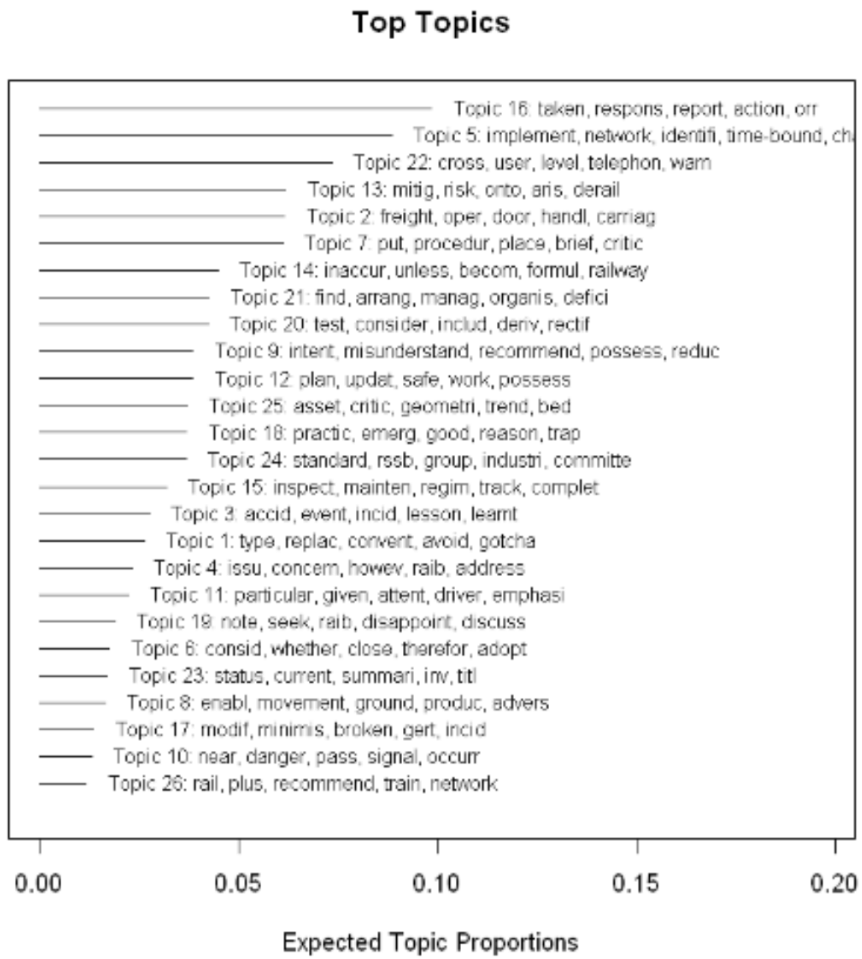


Figure 5-5: Extracted topics and keywords of the RAIB recommendation dataset from the STM

5.3.3 ATSB, Australia – BERTopic model for topics

Figure 5-6 shows the coherence score over different MCS values on the ATSB dataset. The coherence score reaches its peak when the MCS is 10 and decreases as the MCS increases beyond that point. The coherence score fluctuates around 0.58 and continues to decline after the MCS value reaches 30. On the other hand, Figure 5-7 indicates that when MCS is 10, the result shows fewer outliers and smooth distribution of topics. Therefore, the MCS is set to 10 to build the BERTopic model on the ATSB dataset.

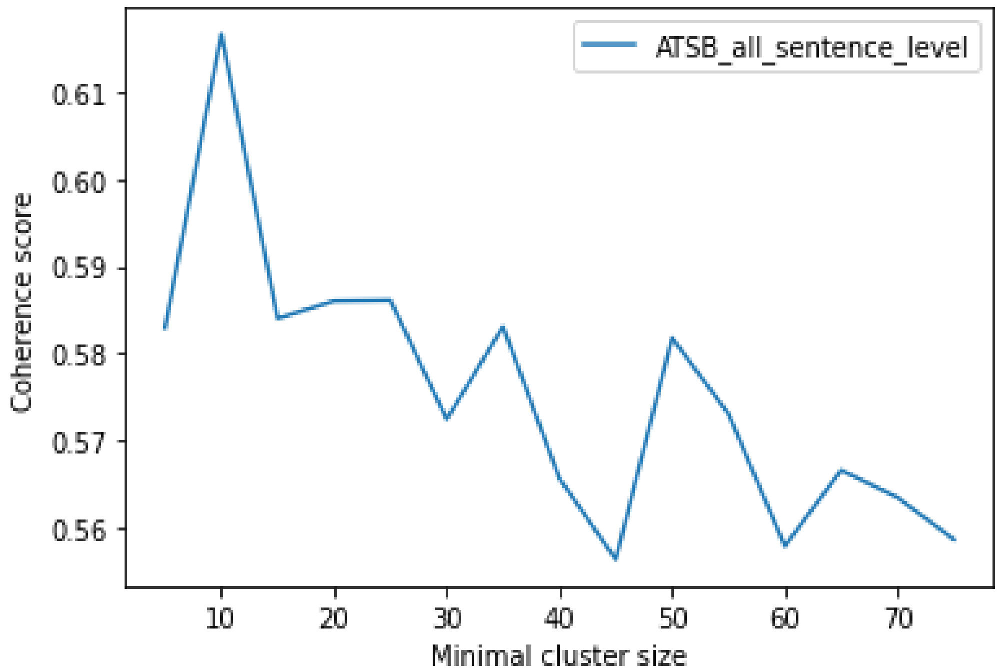


Figure 5-6: Coherence score over different minimum cluster size (MCS) values (ATSB)

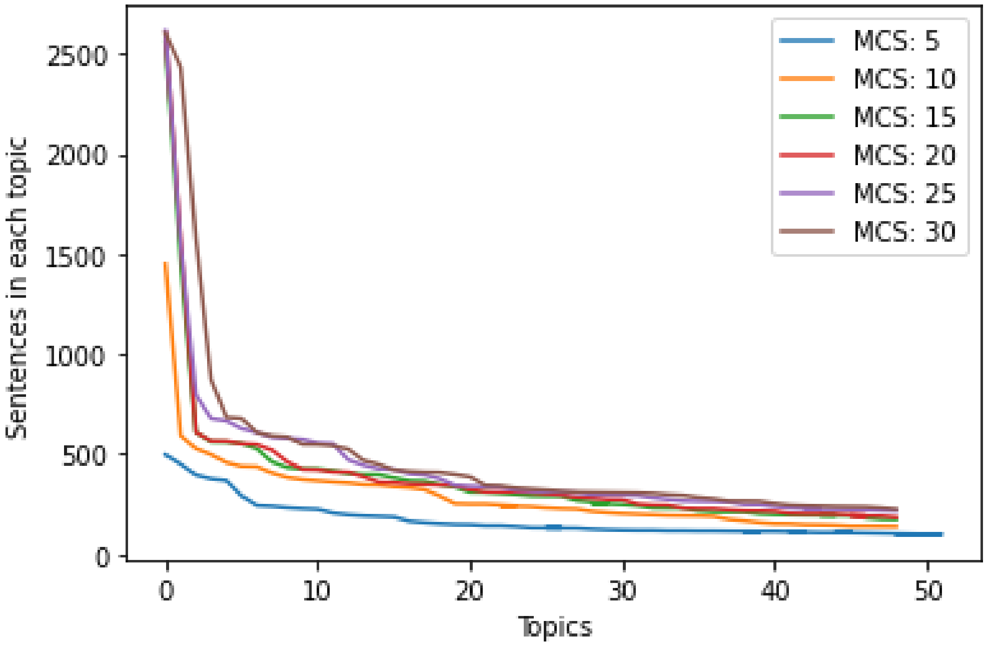


Figure 5-7: Distribution of the number of sentences over each topic with different minimum cluster size (MCS) values (ATSB)

Table 5-4 shows the extracted topics with a high probability of occurrence, associated keywords, and the CV values after removing irrelevant topics. Several distinct topics are identified, such as buckling hazards (topics 6, 11), rail creeps (topics 35, 43, 34), signal conditions (topics 13, 29) and level crossing hazards (topics 37, 39). Figure 5-8 illustrates the inter-topic distance map and indicates several interface issues identified by the investigator, such as the driver–train interface. In addition, it seems that buckling, failure of axles, flange climbing, and rail creep might have significantly impacted the Australian railway system because several relevant topics are identified by the BERTopic model.

Table 5-4: Topic descriptions and coefficient of variance (CV) of ATSB dataset

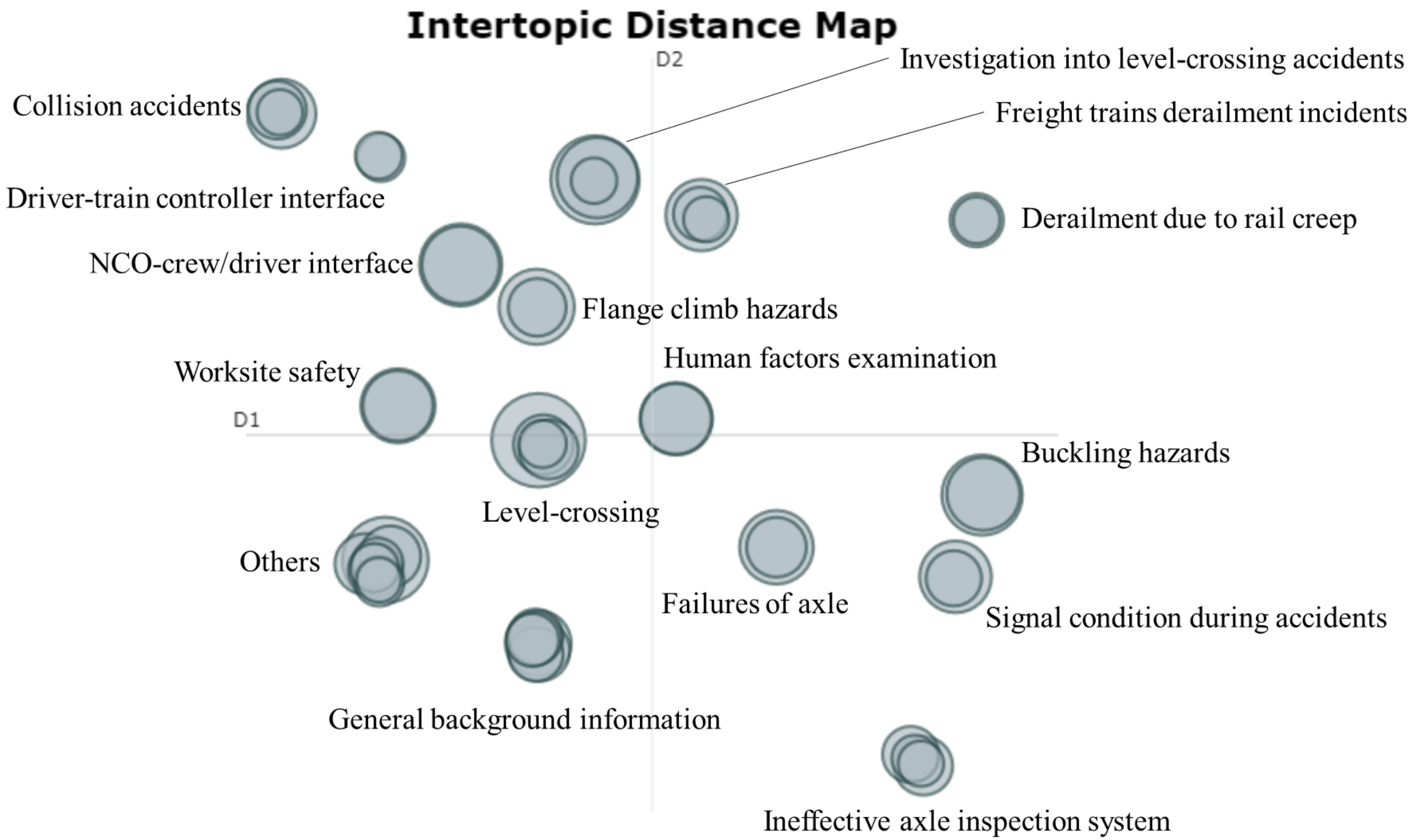
<table>
<tr><th>topic</th><th>Topic – local</th><th>Topic – interval</th><th>Topic – global</th><th>CV</th></tr>
<tr><td>33</td><td>Description of gross mass and containers on wagons</td><td rowspan="2">Background information of incidents</td><td rowspan="4">Freight train derailment incidents</td><td>8.62</td></tr>
<tr><td>46</td><td>Description of train information (length, number of crew…)</td><td>0.73</td></tr>
<tr><td>14</td><td>Consequence of wagons after derailment</td><td rowspan="2">Consequent of derailment</td><td>3.50</td></tr>
<tr><td>48</td><td>Details of bogies' condition during derailment</td><td>5.36</td></tr>
<tr><td>45</td><td colspan="3">Track infrastructure details</td><td>4.25</td></tr>
<tr><td>17</td><td colspan="3">Organisations receiving the draft of the accident report</td><td>0.54</td></tr>
<tr><td>6</td><td>Conditions of ballast crib and shoulder</td><td rowspan="2">Buckling hazards</td><td rowspan="4">Derailment due to buckling hazards and flange climbing</td><td>50.15 (9)</td></tr>
<tr><td>11</td><td>Conditions of sleeper</td><td>43.47</td></tr>
<tr><td>8</td><td>Flange climb accident</td><td rowspan="2">Flange climb hazards</td><td>31.94</td></tr>
<tr><td>27</td><td>The gauge condition prior to accidents</td><td>15.18</td></tr>
<tr><td>44</td><td>Infrastructure maintenance regime and inspection</td><td colspan="2" rowspan="2">Monitoring asset condition via fault monitoring and maintenance regimes</td><td>4.96</td></tr>
<tr><td>47</td><td>Track patrols/ inspection</td><td>5.71</td></tr>
<tr><td>23</td><td colspan="2">Bearing failure</td><td rowspan="3">Failures of axle</td><td>33.46</td></tr>
<tr><td>10</td><td>Conditions of axle bearing</td><td rowspan="2">Ineffective axle inspection system</td><td>118.56 (5)</td></tr>
<tr><td>25</td><td>Defects inspection (continuous ultrasonic testing)</td><td>14.52</td></tr>
<tr><td>40</td><td>Asset Standards Authority (ASA)/ buffer stop</td><td colspan="2" rowspan="3">Asset owner–leaser interface</td><td>35.55</td></tr>
<tr><td>28</td><td>Chicago Freight Car Leasing Australia (CFCLA)/ draft key</td><td>63.93 (8)</td></tr>
<tr><td>36</td><td>Falling jumbo coils</td><td>137.36 (4)</td></tr>
</table>

Table 5-4: Topic descriptions and coefficient of variance (CV) of ATSB dataset (continued)

<table>
<tr><th>topic</th><th>Topic – local</th><th>Topic – interval</th><th>Topic – global</th><th>CV</th></tr>
<tr><td>35</td><td>Rail creep/ monuments</td><td rowspan="2">High temperature hazards to tracks</td><td rowspan="3">Derailment due to rail creep</td><td>20.60</td></tr>
<tr><td>43</td><td>Track temperature</td><td>5.78</td></tr>
<tr><td>34</td><td colspan="2">Determined environmental conditions</td><td>1.22</td></tr>
<tr><td>18</td><td>Conditions of battery cells</td><td colspan="2" rowspan="2">Wire–pantograph interface</td><td>221.72 (1)</td></tr>
<tr><td>21</td><td>Conditions of Overhead Line Equipment (OHLE)/ circuit breaker</td><td>178.10 (2)</td></tr>
<tr><td>2</td><td>Speed of the train</td><td>Conditions of the train</td><td rowspan="5">Investigation into level-crossing accidents</td><td>4.71</td></tr>
<tr><td>5</td><td>Data logger/ Hasler data</td><td>On-board recorders</td><td>6.08</td></tr>
<tr><td>1</td><td>Sounded horn/ audibility</td><td rowspan="2">Events during level crossing incidents</td><td>22.20</td></tr>
<tr><td>22</td><td>Driver behaviour during level crossing</td><td>16.34</td></tr>
<tr><td>24</td><td>Sighting distance/ viewing angle</td><td>Design of level crossing</td><td>10.43</td></tr>
<tr><td>13</td><td>Conditions of signal/ turnout indication/ colour light</td><td colspan="2" rowspan="2">Signal condition during accidents</td><td>32.26</td></tr>
<tr><td>29</td><td>Signal displaying during accident</td><td>23.54</td></tr>
<tr><td>19</td><td colspan="3">Train conditions</td><td>2.29</td></tr>
<tr><td>31</td><td>Consequence of the accident</td><td colspan="2">Fatal/ severe/ minor injuries</td><td>3.71</td></tr>
<tr><td>26</td><td colspan="2">Collision accidents</td><td rowspan="3">Collision between trains on track</td><td>1.72</td></tr>
<tr><td>41</td><td>Description of train’s movement</td><td rowspan="2">Driver–train controller interface</td><td>5.89</td></tr>
<tr><td>42</td><td>Special Proceed Authority (SPA)</td><td>47.22 (10)</td></tr>
</table>

Table 5-4: Topic descriptions and coefficient of variance (CV) of ATSB dataset (continued)

topic	Topic – local	Topic – interval	Topic – global	CV
9	Track Occupancy Authority (TOA)	Worksite safety – worker–train interface	Worksite safety planning – staff, signalling systems, and trains	100.17 (6)
16	Protection Officer (PO) arrangements	Worksite safety – worksite safety planning		21.55
4	Network Control Office (NCO) and crew	NCO–crew/driver interface		24.28
7	Shunt operations			74.91 (7)
37	Australian Level Crossing Assessment Model	Level crossing hazard mitigation strategy		8.15
39	Level crossing safety			20.18
12	Alcohol and drugs tests	Conditions of staff during the rail safety work	Human factors examination	2.55
0	Fatigue investigation	Distraction due to fatigue		28.21
15	Medical examinations and fitness of standards	Medical qualification reviewing		41.82
30	Maintenance of competency (MOS) assessment (training, knowledge gaining for staff)			161.10 (3)
32	V/Line Pty Ltd	Specific organisations mentioned in reports with high frequency		9.94
3	Queensland Rail (QR)			17.86
20	SPAD events due to violation of rules or procedures			29.24



* Note: The x and y dimensions in this diagram are the results of reduced dimensionality with UMAP. Features in higher dimensions are squeezed and only interpretable by computers rather than humans.

Figure 5-8: The inter-topic distance map of identified topics of the ATSB dataset

5.3.4 ATSB, Australia – STM for recommendations data

For the ATSB, making recommendations on investigated railway accidents is selective. Instead, relevant organisations involved and proactive safety actions conducted by operators are thoroughly reviewed before any formal recommendation is made (ATSB, 2011). The ATSB only publishes descriptions of safety issues and proactive actions in place if a recommendation is made, resulting in missing data in the recommendation dataset. Therefore, only 291 recommendations are found in the ATSB dataset, and other reports only disclose accepted proactive actions (n = 191) and safety advisory notices (n = 44). For instance, in the accident report *Level crossing collision between truck and passenger train 8753, Phalps Road, Larpent, Victoria, on 13 July 2016* (ATSB, 2019), the ATSB published the safety issue:

> *"The interaction between V/Line and the Colac Otway Shire Council was ineffective at addressing identified sighting issues at the Phalps Road level crossing."*

Subsequently, proactive actions were received from V/Line Pty Ltd:

> *"V/Line has established a new rail interface team that has been tasked with actively engaging Councils."*

On the other hand, Colac Otway Shire Council also advised the following action:

> *"Colac Otway Shire Council advised that it was committed to its working relationship with V/Line and documenting solutions agreed by each party. The Council advised it would seek to clarify those crossings with outstanding sighting issues and investigate short-to-medium term solutions for implementation."*

The ATSB states that safety actions have been taken to address the safety issue and closed this issue. For the best consistency of input data, this study only uses 291 recommendations for building the STM. Figure 5-9 shows the sematic coherence and exclusivity score over each topic number in the ATSB recommendation dataset. The number of topics is determined as 21 due to the best balance between semantic coherence and exclusivity.

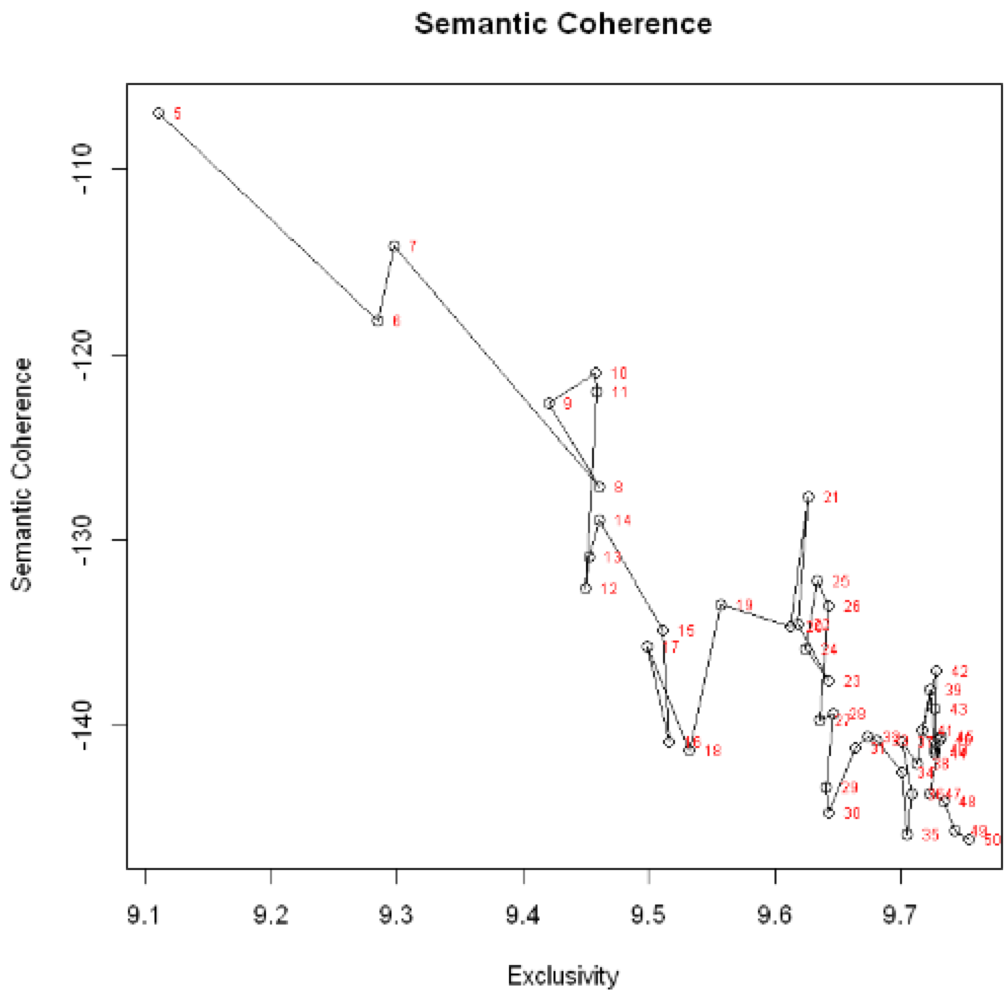


Figure 5-9: Sematic coherence and exclusivity score over each topic (ATSB)

The extracted topics, keywords and assigned topic names are illustrated in Figure 5-10. Frequently proposed recommendations are reviewing communication technology, exchanging knowledge with other organisations, and undertaking risk mitigation strategies. A notable finding is that a more significant proportion of the sentences are sorted to the topic “request to take action to address identified safety issue”, implying that some passive recommendations are made to remind about compliance with existing rules or procedures.

ATSB recommendation dataset

4. Request to take action to address identified safety issues
9. Review communication technologies and cabin design
6. Exchange knowledge with other organisations to reduce level crossing risk
13. Undertake a review for risks mitigation strategies (lesson learnt process)
17. Introduce and update the regulations of freight loading manual
11. Introduce signal protection protocol and review permitted speed
15. Review the effectiveness of testing equipment/ process
5. Develop documented process or specification for safety management system
19. Assess the effectiveness of measurement of monitoring process
16. Develop rules and procedures for governing coordination between organisations
3. Review requirements of track maintenance and monitoring standards
8. Consider appropriate modifications to the assessment and analysis interpretations
18. Assess the effectiveness of maintenance practices and data recording
20. Develop and validate procedures for monitoring management
7. Revise the testing and examining process/ maintenance schedules
14. Introduce framework/ exercises for prioritising tasks during emergency
1. Offer suggestions for instructions of maintenance/ operation process
2. Review process/ regulations of lookout working and worksite safety management
10. Advise on monitoring and reporting shift lengths and break between shifts
21. Ensure compliance to existing regulations on Safeworking arrangements
12. Introduce/ modify the track stability management plan

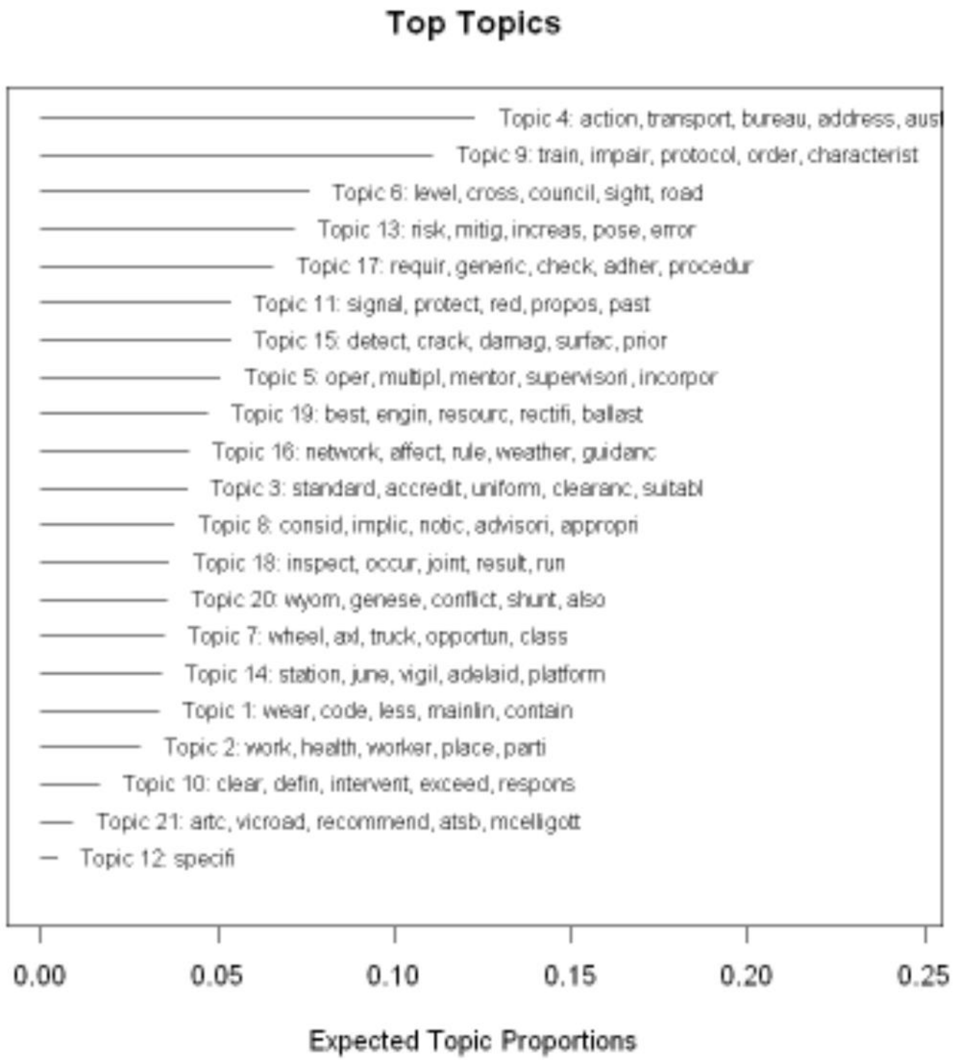


Figure 5-10: Extracted topics and keywords of the ATSB recommendation dataset from the STM

5.3.5 NTSB, US – BERTopic model for topics

For the NTSB dataset, 394 reports published between 1967 to 2022 were retrieved. Of those, 120 reports published before 1995 are scanned documents and were removed, leaving 274 documents to be analysed.

Figure 5-11 shows the NTSB dataset's coherence score over different MCS values. The coherence score reaches its lowest point when the MCS is 25 and peaks when the MCS is 40. The coherence score drops again after the MCS is beyond 40. On the other hand, Figure 5-12 indicates that the number of outliers is around 1,400 when the MCS is below 30 and over 2,000 when the MCS is above 40. Despite a higher number of outliers, this study set the MCS to 40 as a significant improvement of the coherence score on the NTSB dataset.

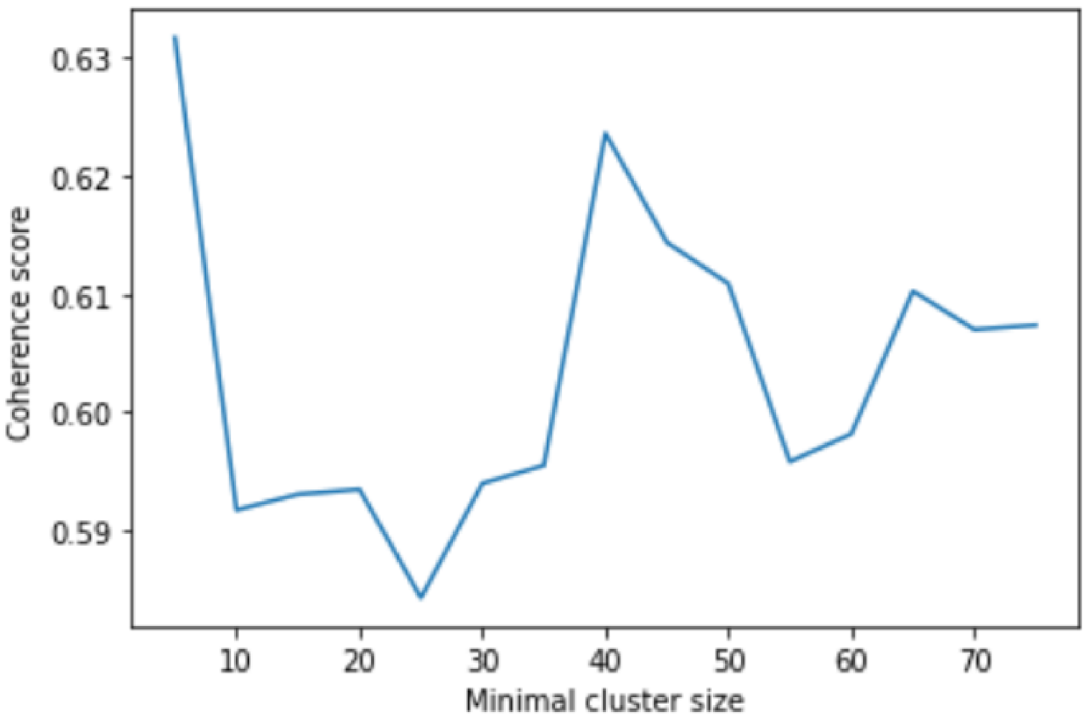


Figure 5-11: Coherence score over different minimum cluster size (MCS) values (NTSB)

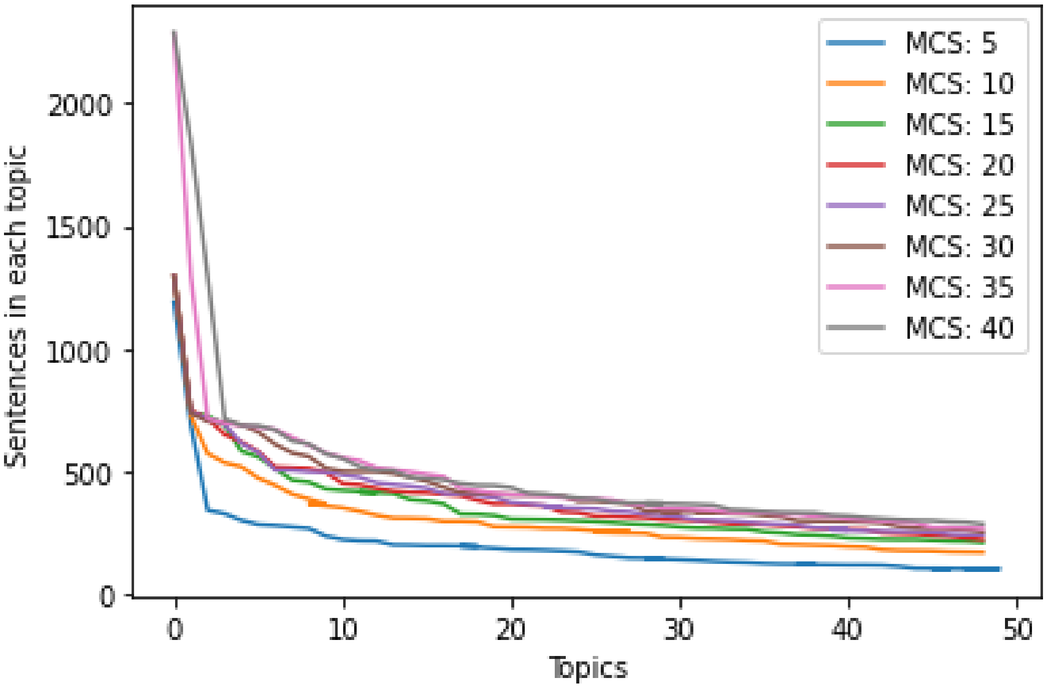


Figure 5-12: Distribution of the number of sentences over each topic with different minimum cluster size (MCS) values (NTSB)

Table 5-5 shows the extracted topics with a high probability of occurrence, associated keywords, and the CV values after removing irrelevant topics in the NTSB dataset. Similar to the ATSB dataset, the issue of the interface between frontline workers is found (topics 43, 9, 27). In addition, accidents relating to freight trains and tank cars (topics 48, 14, 40) are identified and extracted as significant topics as well. The hazard of level crossings plays an essential role in the NTSB dataset with the discussion of types of level crossings (topics 10, 36, 22, 31) and the mechanism of level crossing accidents (topics 29, 37). The emergency response after accidents is also highlighted with high frequency (topic 0). Figure 5-13 illustrates the NTSB dataset's inter-topic distance map of identified topics. Other topics are about the radio communication and the CV value is relatively low, implying that much emphasis is put on crew and staff communication during the investigation. Additionally, some rare elements, such as hazards impacting infrastructure (topics 19, 20) and subway environment control systems (topics 16, 47), are identified by the model, indicating a distinguishing mechanism of railway accidents is found.

Table 5-5: Topic descriptions and coefficient of variance (CV) of NTSB dataset

<table>
<tr><th>topic</th><th>Topic – local</th><th>Topic – interval</th><th>Topic – global</th><th>CV</th></tr>
<tr><td>2</td><td colspan="3">Event, audio and image recorder</td><td>27.784</td></tr>
<tr><td>29</td><td>Occurrence of emergency brake</td><td rowspan="2">Speed at the occurrence of emergency brake</td><td rowspan="10">Grade crossings hazards</td><td>1.982</td></tr>
<tr><td>37</td><td>Speed of the train recorded</td><td>1.346</td></tr>
<tr><td>15</td><td>Condition of the signal aspect</td><td rowspan="2">Issue of grade crossings design</td><td>8.127</td></tr>
<tr><td>49</td><td>Pre-emption/ “all-red-flash” design of grade crossings</td><td>69.527</td></tr>
<tr><td>10</td><td>High-risk private highway–railroad grade crossings</td><td rowspan="2">Hazard of private highway–railroad grade crossings</td><td>37.728</td></tr>
<tr><td>36</td><td>Bus driver training about grade crossing in school district</td><td>126.415</td></tr>
<tr><td>22</td><td>Sounded horn/ audibility</td><td rowspan="2">General grade crossings hazards</td><td>9.221</td></tr>
<tr><td>31</td><td>Hazard of stopping within the boundary of the crossing</td><td>24.689</td></tr>
<tr><td>28</td><td>Consequence of derailment</td><td colspan="2" rowspan="2">Condition of the train</td><td>2.270</td></tr>
<tr><td>44</td><td>Components/ units of the train</td><td>2.718</td></tr>
<tr><td>16</td><td>Conditions of the tunnel ventilation system</td><td rowspan="2">Subway environment control system</td><td rowspan="2">Subway environmental hazards</td><td>263.979</td></tr>
<tr><td>47</td><td>Electrical arcing due to water intrusion</td><td>65.671</td></tr>
<tr><td>24</td><td>Pressure of brake/ relief valve</td><td rowspan="2">Failure of brake system</td><td rowspan="2">On-board equipment hazards</td><td>53.505</td></tr>
<tr><td>38</td><td>Air leakage from the brake pipe</td><td>27.027</td></tr>
</table>

Table 5-5: Topic descriptions and coefficient of variance (CV) of NTSB dataset (continued)

topic	Topic – local	Topic – interval	Topic – global	CV
6	Working conditions of staff			9.596
18	Weather conditions			6.019
48	Unsafe offloading practices of solvent blend wastes		Tank cars hazards	141.007
14	Cracked or broken joint bars/ bolts	Tank cars failure and certifications		42.103
40	Specifications for tank cars			83.276
30	Damages to assets			2.297
19	Bridges' capacity to carry floods	Hazards of bridges	Infrastructure hazards	254.145
20	Escorting permit loads			231.054
8	Parasitic oscillation of track circuit modules	Failures of trains' circuits		474.935
25	Failures of emergency windows/ doors			29.753
11	Organisational culture of safety oversight	Rail safety oversight framework	Worksite hazards	115.046
42	Regulation of State oversight agency			27.552
3	Installation of Positive Train Control			22.892
17	Unsafe work practices culture of Amtrak's management	Safety culture		82.881
33	Safety Management Manual and safety culture			21.074

Table 5-5: Topic descriptions and coefficient of variance (CV) of NTSB dataset (continued)

topic	Topic – local	Topic – interval	Topic – global	CV
46	Operating rules for employees	Requirements of employees’ conditions	Hazards of employees’ medical conditions	12.903
1	Drug, alcohol, and toxicology test			20.740
32	Efficiency of tests			3.610
23	Colour vision test	Medical conditions		56.817
39	Obstructive sleep apnoea			19.730
13	Conditions of track inspections		Switches and tracks hazards	26.078
4	Conditions of switches	Failure of switches		46.900
35	Subdivision of tracks			1.607
12	Fatalities and injuries			3.544
26	Interface between conductors and railroad cars	Conductors’ failure		3.764
41	Conditions of conductors			2.433
0	Emergency response after accidents	Emergency response of train operators		53.744
21	Operation of CSX Transportation and MARC Train			53.055
43	Track warrant authority (interface between train crews and the dispatchers)	Interface between train crews and the dispatchers		50.307
9	Usage of cell phones and text messages			46.934
27	Radio communications between crew members and dispatchers			3.936

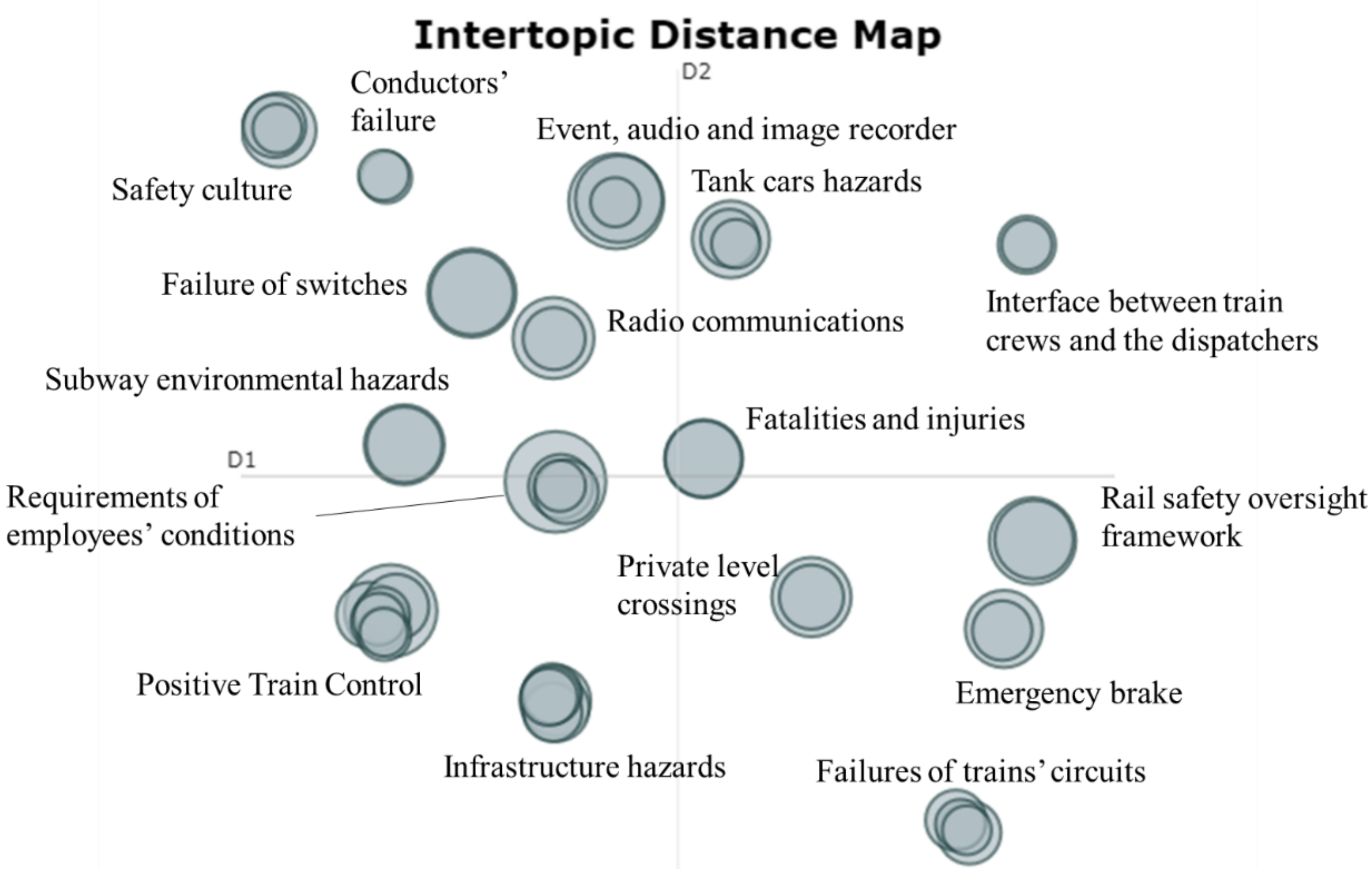


Figure 5-13: The inter-topic distance map of identified topics of the NTSB dataset

5.3.6 NTSB, US – STM for recommendations data

Figure 5-14 shows the sematic coherence and exclusivity score over each topic number in the NTSB recommendation dataset. The number of topics is set to 12 for the best combination of sematic coherence and exclusivity score. Figure 5-15 illustrates the extracted topics and keywords of the NTSB recommendation dataset from the STM. A considerable amount of focus is put on cooperation with organisations within the railway industry, implying less intervention and restrictions on the approach operators apply to address identified hazards. Furthermore, assisting research and programs is also mentioned in high frequency, which might indicate the promotion of cooperating with third parties and producing a comprehensive solution. Another note is that assigning specific methods to address identified hazards is rarely found in the NTSB recommendations and most of them are supportive of the railway industry.

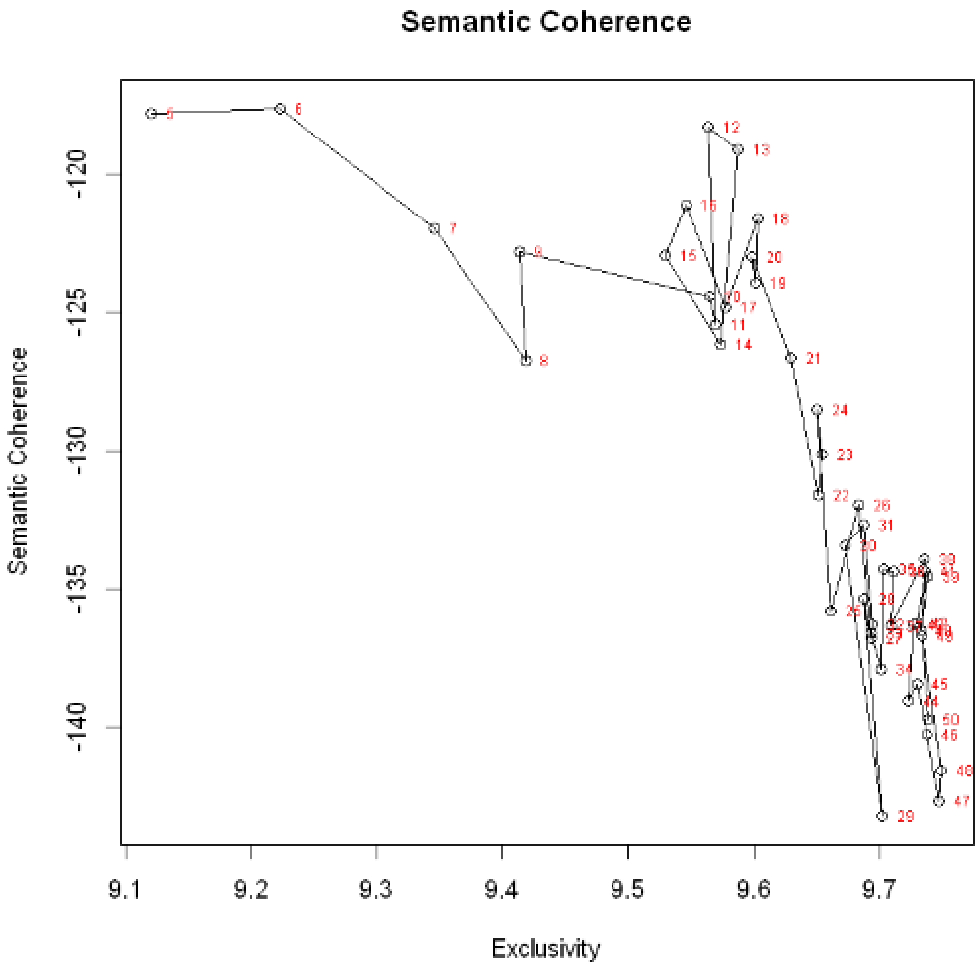


Figure 5-14: Sematic coherence and exclusivity score over each topic number (NTSB)

NTSB recommendation dataset

10. Assist research and programs in establishment of working group
5. Cooperate with other organisations on examination, design and maintenance procedure
8. Cooperate with local train authority to introduce technology/ advanced train control system
4. Verify the existing systems meet industrial standards
12. Develop, implement and update procedures of hazardous material transportation with other organisations
3. Establish supervisory procedures to monitor activities of employees
6. Revise or amend procedures, instructions and regulations
11. Equip/ modify on-board devices i.e., radio
7. Install physical signs at grade crossings
9. Improve visibility/ equipment for emergency conditions
1. Review, develop and implement risk management program for identified deficiencies
2. Disseminate the circumstances of the accident to membership
13. Cooperate with local transit authority on amendment of regulations

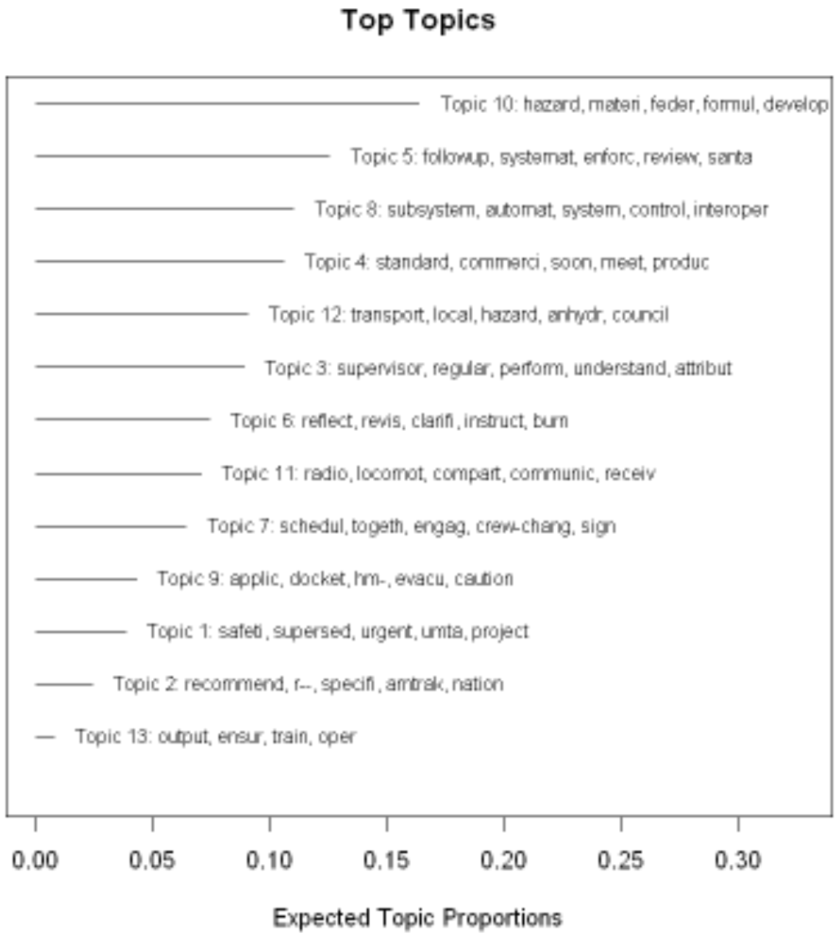


Figure 5-15: Extracted topics and keywords of the NTSB recommendation dataset from the STM

5.3.7 TSB, Canada – BERTopic model for topics

Figure 5-16 shows the TSB dataset's coherence score over different MCS values. The coherence score increases dramatically and reaches a plateau when the MCS is 30. Subsequently, the improvement in the coherence score is limited after the MCS is beyond 35. On the other hand, Figure 5-17 shows that the number of outliers increases significantly after the MCS is beyond 30. Therefore, the MCS is set to 30 to build the model on the TSB dataset.

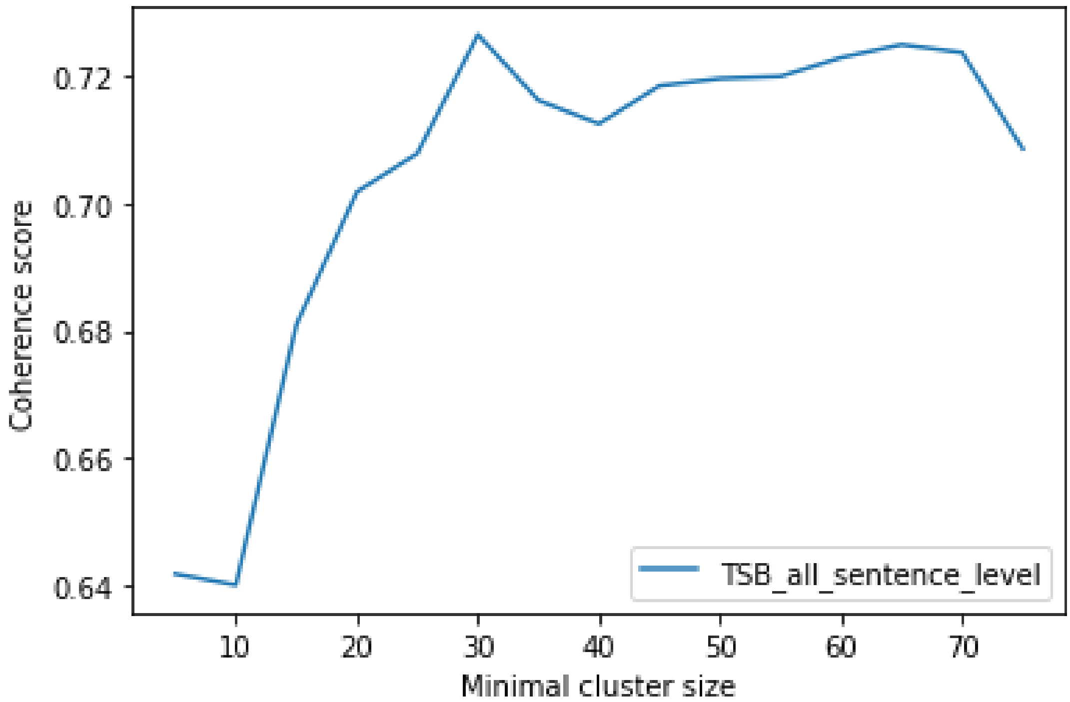


Figure 5-16: Coherence score over different minimum cluster size (MCS) values (TSB)

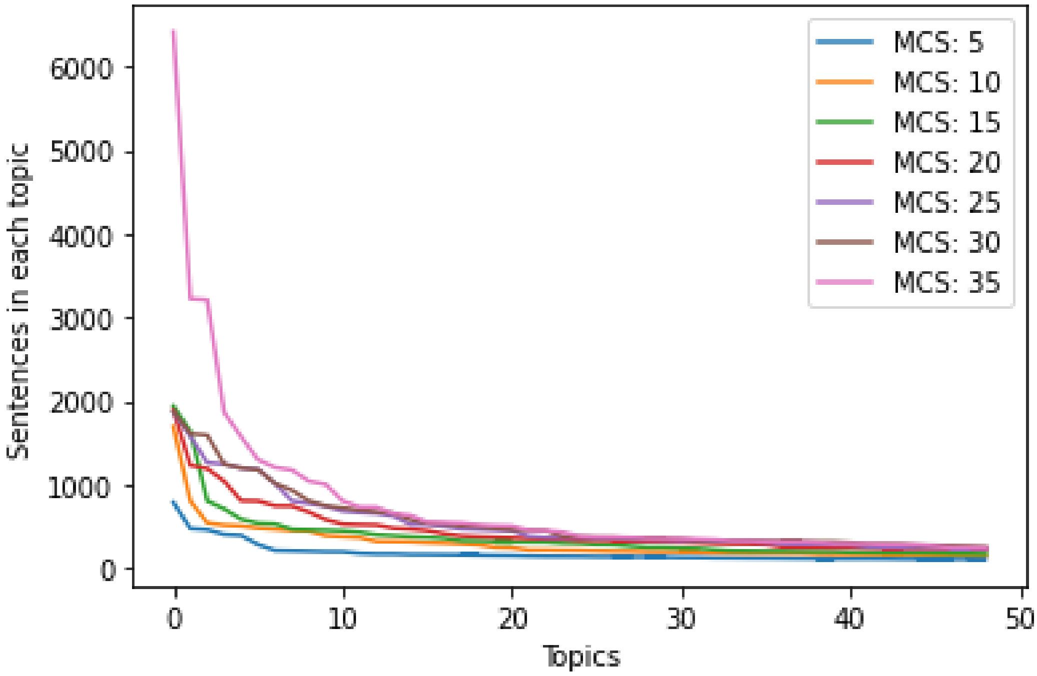


Figure 5-17: Distribution of the number of sentences over each topic with different minimum cluster size (MCS) values (TSB)

Table 5-6 describes the extracted topics with a high probability of occurrence, associated keywords, and the CV values after removing irrelevant topics in the TSB dataset. The topic relating to the interface between crew and staff (topics 4, 17, 19, 25) is identified as in the results of other datasets. The difference between the concept of “human fatigue” (topic 0) and “machine fatigue” (topic 46) is identified and discriminated by BERTopic, indicating the superiority of using the word embedding method compared to the bag-of-words approach. Some specific topics relevant to the hazards to freight trains are also revealed, such as excessive truck shunting (topic 27) and the risk of flammable materials (topic 18), which might indicate the potential risks of the development of the freight rail industry. Figure 5-18 illustrates the TSB dataset’s inter-topic distance map of identified topics. Topics relating to the derailment of freight trains have been clustered into a dense group on the left-hand side. Another strong topic “safety management system” is clustered as well, indicating the intention of developing a systematic railway safety management. Notably, the hazard of worksite safety is not identified by BERTopic, indicating the relatively low discussion of this topic in the TSB investigations.

Table 5-6: Topic descriptions and coefficient of variance (CV) of TSB dataset

topic	Topic – local	Topic – interval	Topic – global	CV
46	Fracture surface due to fatigue	Rail fracture surface hazards	Rail fracture hazard	2.931
48	Rail fracture			2.629
9	Condition of tie plates and secured spikes	Overview of track information		8.048
38	Track information			0.340
1	Emergency brake application	Occurrence of emergency brake		18.540
11	Brake pipe pressure			74.668
21	Conditions of air brake tests	Brake tests before departing	CROR on special instructions	45.119
28	Certified car inspector			3.083
10	Application of hand brakes	CROR on brake and movement		77.484
26	Canadian Rail Operating Rules (CROR)			2.360
37	Yard assignment description	Interface between workers in the yard		17.613
43	Failure of transfer between yardmasters			16.337
3	Derailment of freight cars	Derailment of freight trains		3.449
42	Location where locomotive came rest			0.909
6	Malfunction of switches	Defects on rail tracks		50.734
34	Damage on tracks			0.793
14	L/V ratios of single-wheel	Measurement on rail wheel		17.482
36	Observation of wheel flange marks			1.261

Table 5-6: Topic descriptions and coefficient of variance (CV) of TSB dataset (continued)

topic	Topic – local	Topic – interval	Topic – global	CV
22	Wheel Impact Load Detector (WILD)	Wheel overloading hazards	Wheel–bearing system interface	28.267
32	Risks associated with Transcona wheel shop loose wheels			74.013
27	Excessive truck hunting/ Constant Contact Side Bearings			46.425
35	Alert for roller bearing temperature			27.467
8	Sounded horn/ audibility		Grade crossings hazards	27.738
2	Warning devices/ rules of grade crossing	Drivers' interface at grade crossings		44.698
13	Behaviour of grade crossing users (driver)			8.368
4	Interface between Rail Traffic Controllers and crew members	Interface between RTC and others	Communications hazards	54.549
17	Interface between foreman and RTC			35.120
19	Display of indication signals			17.111
25	Radio communications			5.056
40	Hazards related to train marshalling			14.556
24	Absence of on-board voice recorders			126.765
7	Emergency response	Hazards of poor design of emergency exits and response		19.017
16	Failure of emergency exit (on board)			54.740

Table 5-6: Topic descriptions and coefficient of variance (CV) of TSB dataset (continued)

<table>
<tr><th>topic</th><th>Topic – local</th><th>Topic – interval</th><th>Topic – global</th><th>CV</th></tr>
<tr><td>5</td><td colspan="3">Flood/ drainage system</td><td>59.779</td></tr>
<tr><td>39</td><td colspan="3">Failure of the thermite weld</td><td>23.253</td></tr>
<tr><td>0</td><td>Fatigue</td><td colspan="2" rowspan="2">Hazards of fatigue</td><td>32.543</td></tr>
<tr><td>31</td><td>Risk of memory lapse</td><td>4.147</td></tr>
<tr><td>18</td><td colspan="3">Risk of flammable materials</td><td>15.346</td></tr>
<tr><td>20</td><td colspan="3">Safety management system</td><td>53.215</td></tr>
</table>

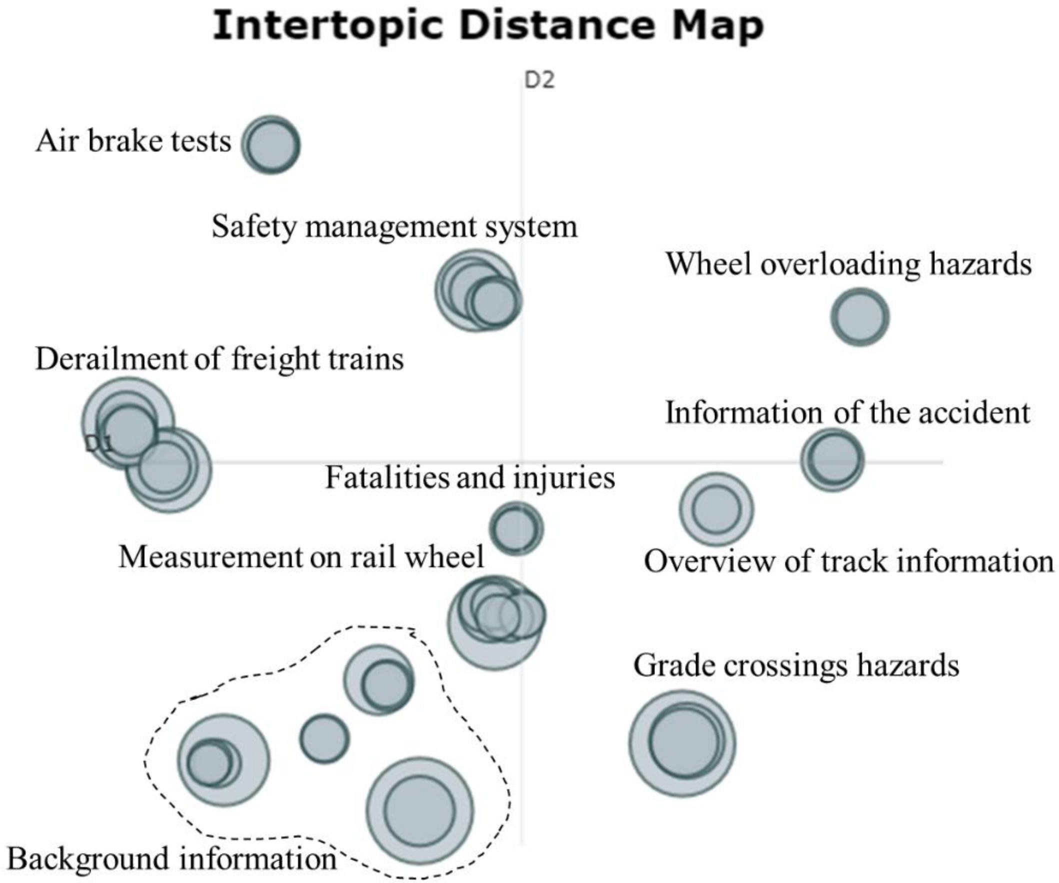


Figure 5-18: The inter-topic distance map of identified topics of the TSB dataset

5.3.8 TSB, Canada – STM for recommendations data

The railway accident investigations conducted by the TSB are classified into six classes according to their relative importance, complexity, and potential for yielding safety lessons (TSB, 2022). Each class has different complexity, level of investigation, processes, and target timeline. For instance, class 1 is a safety issue-related investigation with common features that have formed a pattern over a period and raise or might raise significant risks based on the result of statistical analysis. This type of investigation takes about 730 days (2 years) before the finalisation of the report. Another example is class 4, a limited-scope investigation of accidents resulting in significant consequences, attracting the public's attention but having limited learning points. In this case, no finding or recommendation will be proposed. Notably, only class 2 investigations are required to result in recommendations given that making recommendations is optional in other types of investigations.

To better understand the nature of hazards in Canada, railway accident reports classified as class 1 to 4 are collected, starting from 1991 to 2022 (available period from the TSB website). 415 reports and 73 recommendations (exclusive of the same recommendation for different cases) were retrieved.

Figure 5-19 shows the sematic coherence and exclusivity score over each of topic number on the TSB recommendation dataset. Due to the limited TSB recommendation data, the curve cannot reflect the real performance of the STM given that the model may cluster each input data to an individual topic and result in high semantic coherence and exclusivity. To address this issue, the number of topics is set to 5 to ensure sufficient sentences are sorted to each topic.

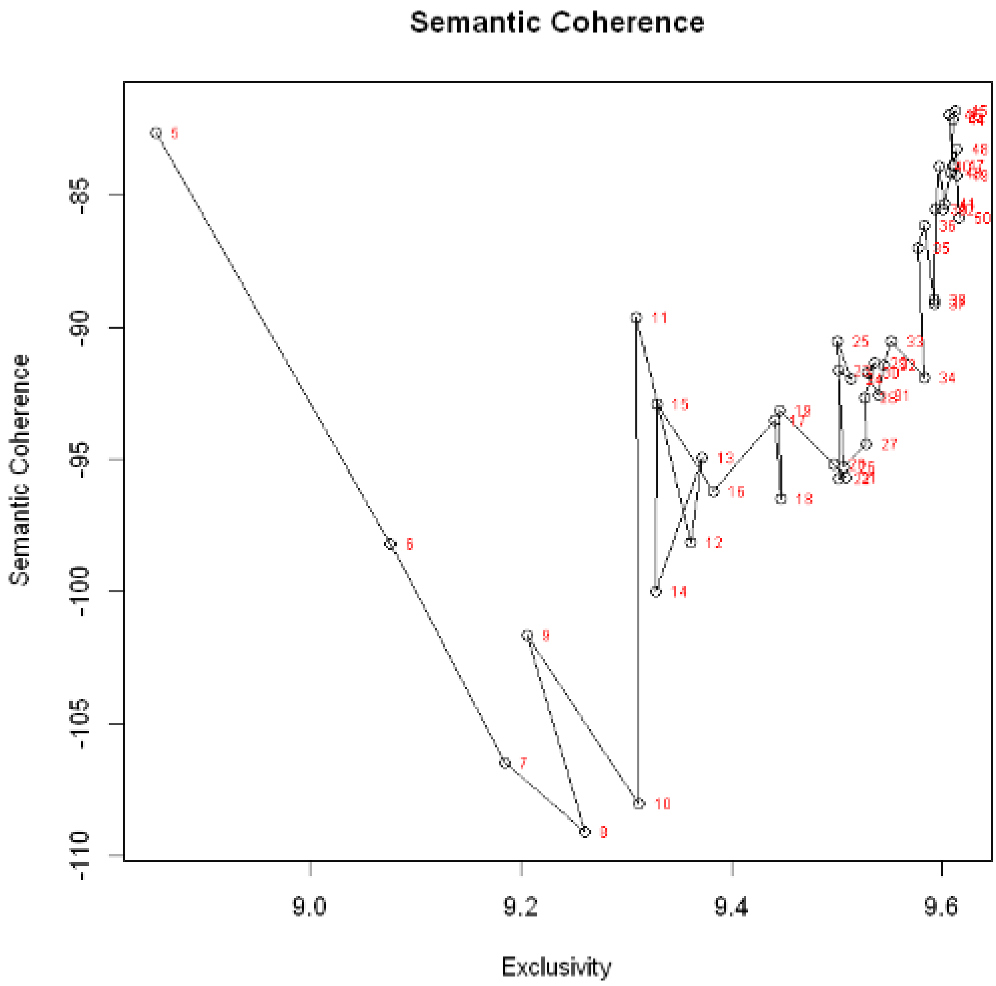


Figure 5-19: Sematic coherence and exclusivity score over each topic (TSB)

Figure 5-20 illustrates the extracted topics and keywords of the TSB recommendation dataset from the STM. The TSB dataset comprises recommendations requesting the examination and reassessment of current procedures rather than developing new rules or processes. On the other hand, limited suggestions are given to cooperate with organisations within the railway industry. Furthermore, most recommendations are directed to individual railway companies, assigning an objective to resolve identified hazards. Last, interfering recommendations such as assigning specific methods to organisations involved are not identified, implying the TSB tends to primarily propose supportive advice and retains a large degree of flexibility for the railway industry.

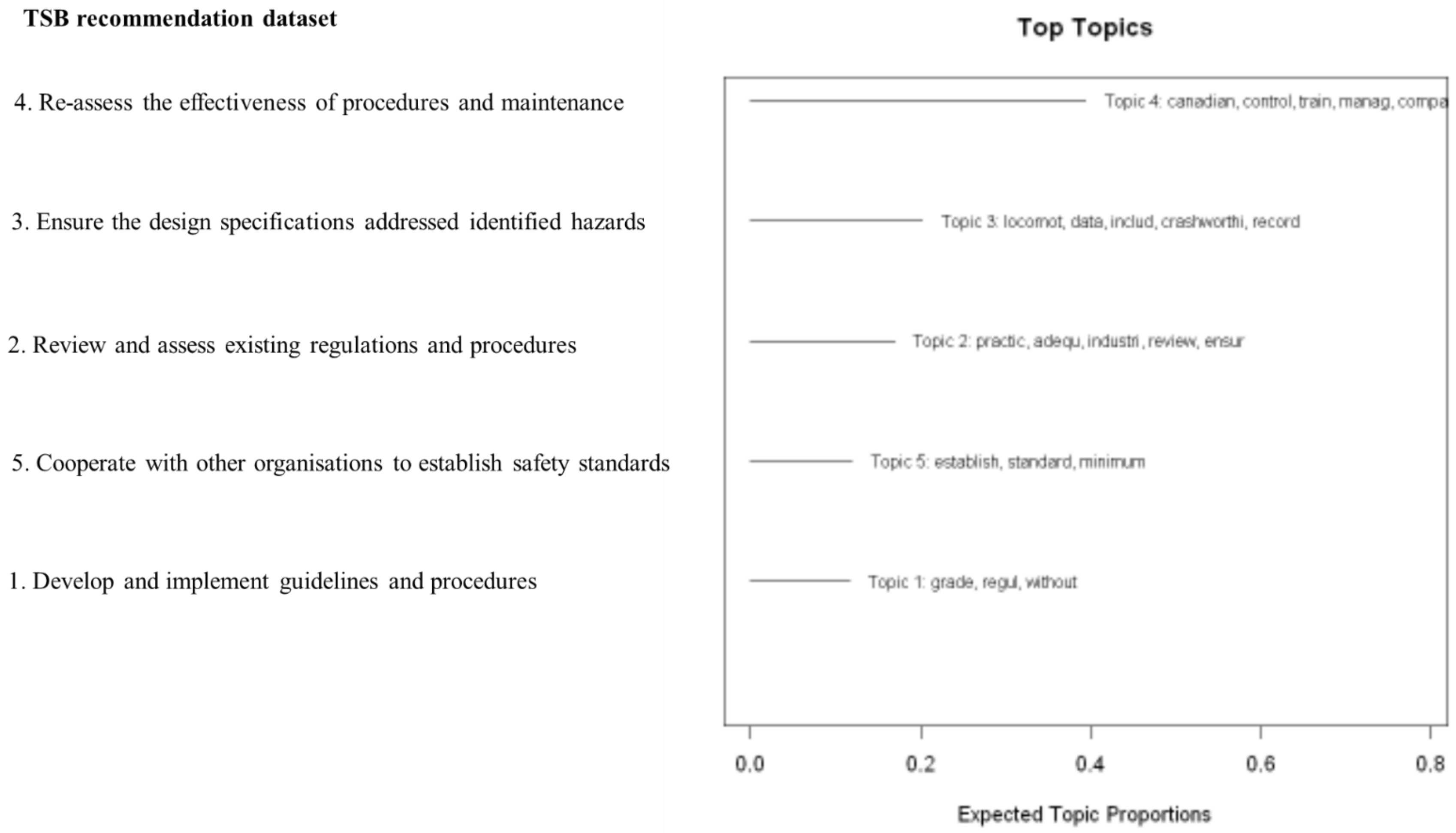


Figure 5-20: Extracted topics and keywords of the TSB recommendation dataset from the STM

5.4 The ontology, knowledge graph selection and entity linking

To address the interface between railway accident investigation reports, Wikidata and railway accident ontology, the Wikidata Graph Builder is introduced to determine the connections between entities. The Wikidata Graph Builder is a tool visualising the construction of an item and its properties on the Wikidata ontology (Nielsen, 2019). The input items and properties can be any entity on Wikidata, and the Wikidata Graph Builder builds a class tree with the input item at the centre and the property as edges. For example, Figure 5-21 illustrates the class tree generated by the Wikidata Graph Builder with the input entity "rail freight company" (the blue dot) and property "subclass of" (edges). The entity "rail freight transport" is the subclass of the concept "rail transport" and "land transport and transport via pipelines" at the higher hierarchy of the Wikidata ontology.

On the other hand, entities that are the subclass of "rail freight transport" are also extracted, such as the entity "double-stack rail transport". The class tree is further extended by extracting entities having the property "subclass of" of the identified concepts until the root entity. Note that each concept or entity listed on the class tree contains a series of instances not included in the diagram for a concise view. For example, entities "Bowmans Rail", "One Rail Australia" and "SCT Logistics" can be found as instances of the concept "rail freight transport" in the Wikidata ontology.

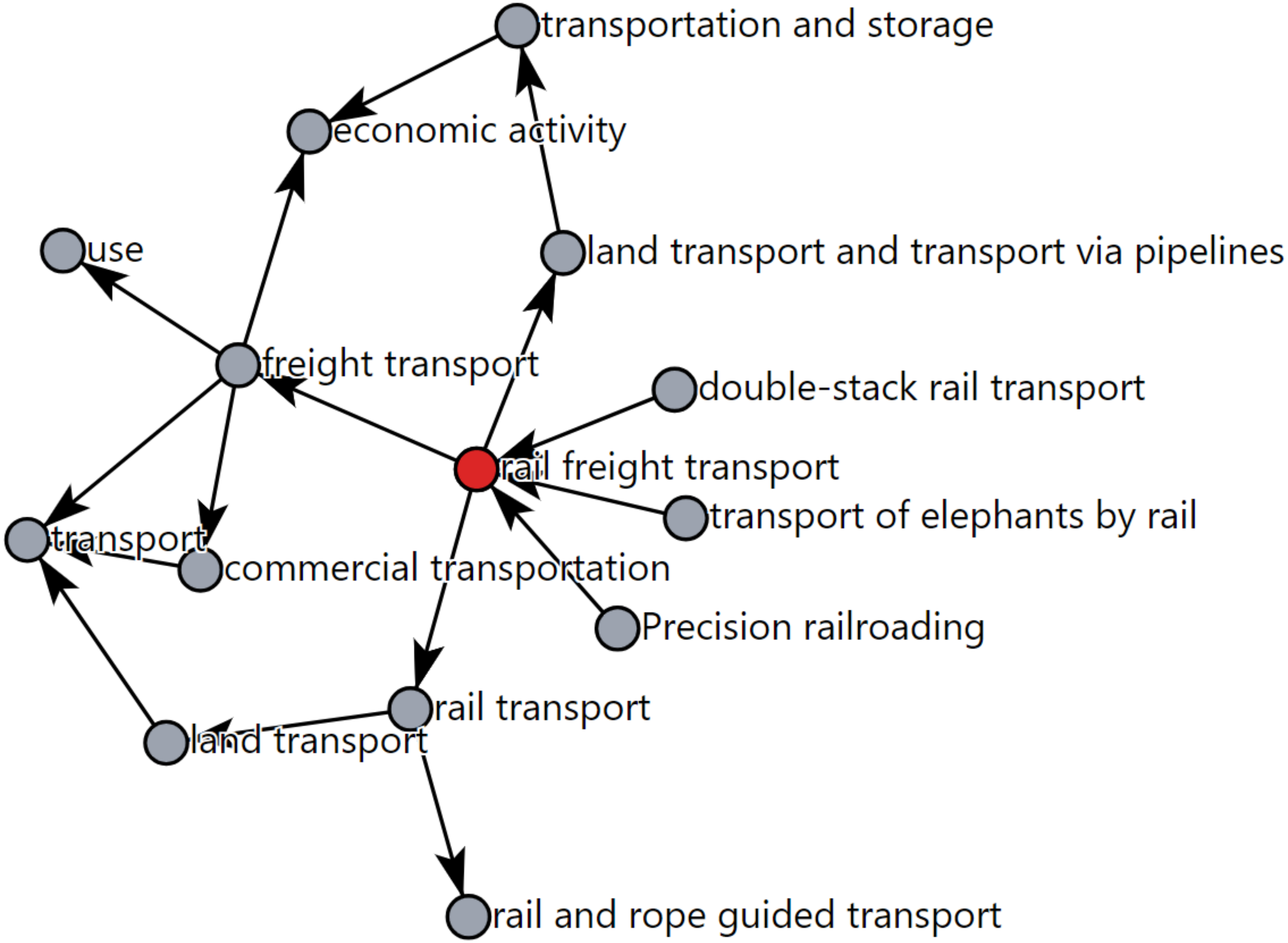


Figure 5-21: Class tree generated by the Wikidata Graph Builder with the entity "rail freight company" and property "subclass of"

Subsequently, the designed railway accident ontology is further extended by connecting relevant entities to concepts in the Wikidata ontology. The linked concepts are carefully selected to meet the scale of the description. For instance, the “transport company” concept from the Wikidata ontology is selected to link to our entity “organisation”. However, the entity “train operating company” might be feasible to be used as well considering that the railway industry constitutes a wide range of organisations such as rolling stockholders and rail vehicle manufacturers, a concept at a higher level might be appropriate to cover all entities of interest.

On the other hand, several properties are available in the Wikidata ontology. However, only the property “instance of” is applied to connect entities identified by the *Tagme* tool because it reveals the hierarchy structure of the ontology. Table 5-7 shows railway companies identified by *Tagme* from the ATSB dataset. The issue of incorrectly identifying certain entities with the same name, such as the Southern Railway in the UK in Table 5-7, could be addressed by using the context-sensitive disambiguation process (Section 4.1.2.6). Additionally, at least one corresponding Wikidata page must be connected to the identified entity because *Tagme* is established based on the Wikidata SPARQL query service. As a result, the original data is linked to the Wikidata ontology and railway accident ontology. Figure 5-22 illustrates the extended designed railway accident ontology connecting to the Wikidata ontology. Grey boxes with solid outlines indicate that the designed railway accident ontology concept is connected to the Wikidata ontology. In contrast, grey boxes with dotted outlines refer to the connected concepts in the Wikidata ontology with the Q-value identification. Note that not all concepts are linked to the Wikidata ontology but only concepts with overlapped definitions are connected. Thus, the connected ontology can be applied to address the difficulty of various railway terminology used in different countries.

Table 5-7: Extracted railway company from ATSB dataset

	Property (wdLabel)	Class (ps_Label)	Entity (from)
0	instance of	railway company	Line
1	instance of	railway company	SCT_Logistics
2	instance of	railway company	NSW_TrainLink
3	instance of	railway company	Aurizon
4	instance of	railway company	Queensland_Rail
5	instance of	railway company	Pacific_National
6	instance of	railway company	Sydney_Trains
7	instance of	railway infrastructure manager	Australian_Rail_Track_Corporation
8	instance of	railway company	Atchison,_Topeka_and_Santa_Fe_Railway
9	instance of	railway company	Metro_Trains_Melbourne
10	instance of	railway company	TasRail
11	instance of	railway company	Genesee_&_Wyoming_Australia
12	instance of	railway company	AN_Tasrail
13	instance of	railway company	FreightLink

Table 5-7: Extracted railway company from ATSB dataset (continued)

	Property (wdLabel)	Class (ps_Label)	Entity (from)
14	instance of	railway company	Buenos_Aires_Great_Southern_Railway
15	instance of	operator of last resort	Northern_(train_operating_company)
16	instance of	train operating company	Northern_(train_operating_company)
17	instance of	train operating company	Connex_Melbourne
18	instance of	railway infrastructure manager	Railtrack
19	instance of	railway company	Southern_Railway_(UK)

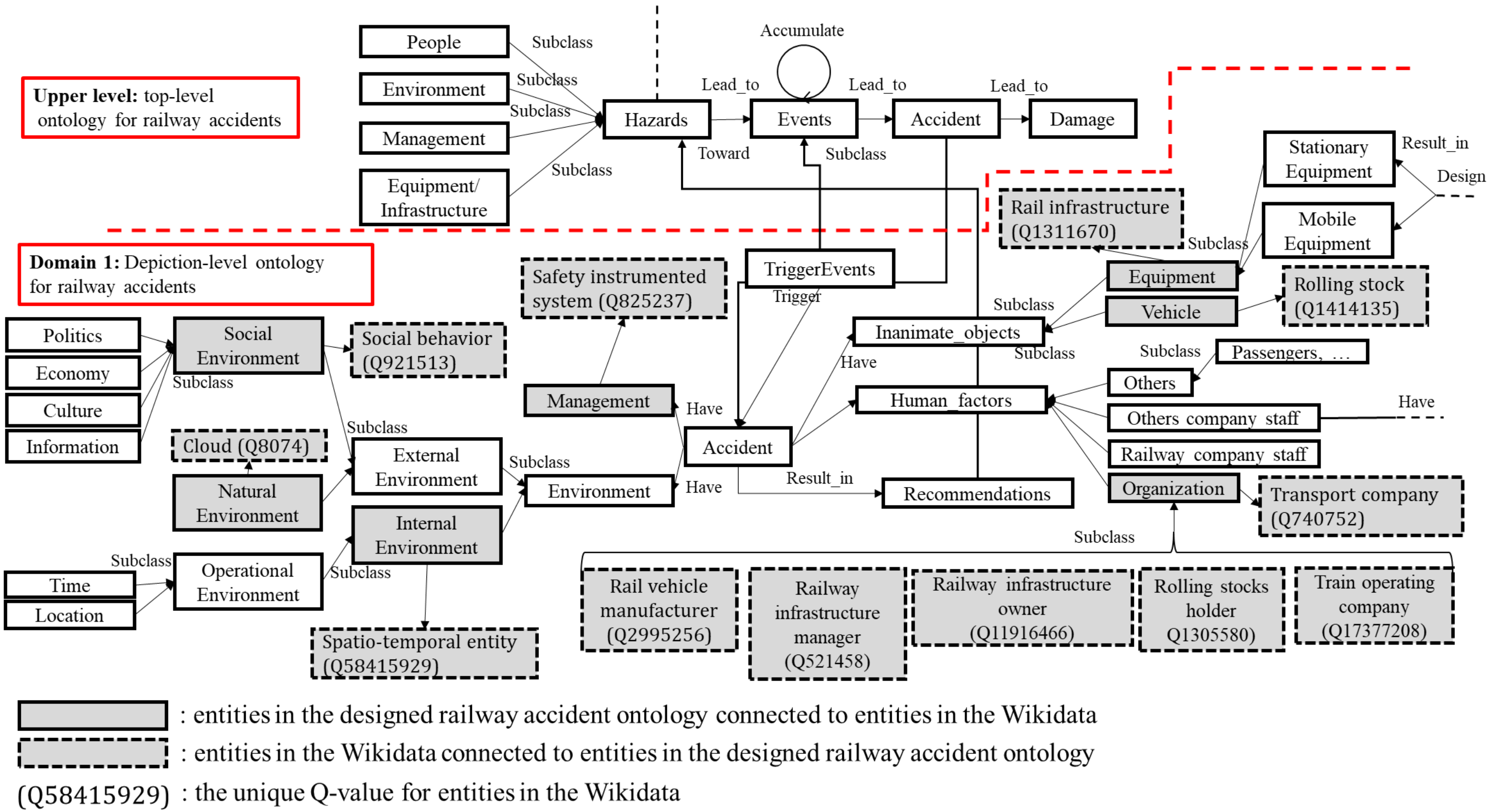


Figure 5-22: The extended designed railway accident ontology connecting to the Wikidata ontology

5.5 Cross-country analysis

To understand the difference of the nature of railway accidents between countries, the cross-country analysis is conducted based on the outcome of each STM and BERTopic model.

5.5.1 Cross-sectional analysis – railway accidents

First, topics from investigated countries with high possibilities of occurrence as discussed in previous sections are extracted and reviewed. Second, keywords under each topic are linked to the designed railway accident ontology and keywords under topics from investigated countries with mentions (original keywords in the text) connected to the same entities are highlighted. These topics, mentions and connected entities are reviewed manually, and similar topics are further labelled as having a connection to indicate the overlapped elements found in railway accident investigation reports of two countries. Other topics without being connected are labelled as country-specific topics. Last, a comparison between topics discussed by each investigation body can be mapped.

Figure 5-8 shows an example of the comparison between identified topics from railway accident reports published by ATSB and RAIB. Some potential differences in the railway accident mechanism can be observed. For instance, in the ATSB dataset buckling hazards, high temperature hazards to tracks and rail creep/monuments hazards are highlighted along with the jumbo coils, implying the impact of heavy freight trains and the high temperature on tracks. In contrast, in the RAIB dataset there is wide discussion of sanding and adhesion issues and holdfast panel–sleeper interface, which might indicate the hazard of the interface between wheels and tracks due to the extreme weather conditions or heavy autumn leaf fall disruption in a short space of time. Nevertheless, the highlighted topics of platform–tram interface/striking pedestrian and trams also suggest a higher level of discussion in the UK on passenger train accidents rather than freight train accidents. This case has shown different fundamental issues that the two countries of Australia and the UK encounter.

Despite some differences in hazards found between countries, a holistic view of the mechanisms that each hazard interacts with others and results in a railway accident is absent. For instance, buckling hazards, high temperature hazards and jumbo coils seem to have a causal relationship to cause a railway accident. However, the extent to which each hazard contributes to a railway accident is not identified yet. Nevertheless, only topics with high possibilities of occurrence are extracted and illustrated in Table 5-8. Other minor but critical elements might be overlooked by only analysing topics with high occurrences given that each railway accident is triggered by different causes (Cozzani et al., 2004). Figure 5-23 shows the example of a joint hazard “groundwater” combined with other hazards in different countries. Each country experienced railway accidents due to this hazard in different ways and addressed it with various approaches. For example, the RAIB in the UK takes the groundwater as an external circumstance and controls the risk by improving the drainage assets. In contrast, the ATSB in Australia investigates the structure of railway tracks and introduces temporary speed restrictions to the influenced sections. As a result, topics shown in Figure 5-23 are identified as having a lower probability of occurrences and being ignored from the outcomes although the potential connections to the mechanism of the railway accident might be substantial. Therefore, a systematic view of hazards is required to address this issue and ensure a comprehensive map of the mechanisms of railway accidents. More details are introduced in Sections 6.2 and 6.3.

Table 5-8: Comparison between identified topics from railway accident reports published by ATSB and RAIB

ATSB (Australia) only	**RAIB (UK) only**	**Common to ATSB and RAIB**
• Conditions of battery cells • Description of gross mass and containers on wagons • Buckling hazards • Asset owner–leaser interface • Jumbo coils • High temperature hazards to tracks • Rail creep/ monuments • Data logger/ Hasler data • Driver behaviour during level crossing • Sighting distance/ viewing angle • Conditions of signal/ turnout indication/ colour light • Signal displaying during accident • Medical examinations and fitness of standards • SPAD events due to violation of rules or procedures	• CCTV, monitor/ recording • Sanding/adhesion • AWS isolation active • Platform–tram interface/ striking pedestrian • Trams • Bolts failure • Stretcher bar • Holdfast panel–sleeper interface • Obstacle detection of doors • Natural hazards (landslip, flood…) • Conditions of engineering units • Earthworks	• Speed • Emergency brake • Sounded horn • Train horn • Site lookout • Fatigue • Communication – Signaller (radio, GMR-S, etc) • Driver knowledge, training, instruction • Cause and result of flange climbing • Grinding repairs • Track defects inspection • Wire–pantograph interface • Conditions of Overhead Line Equipment (OHLE)/ circuit breaker • Failure mode of the axle

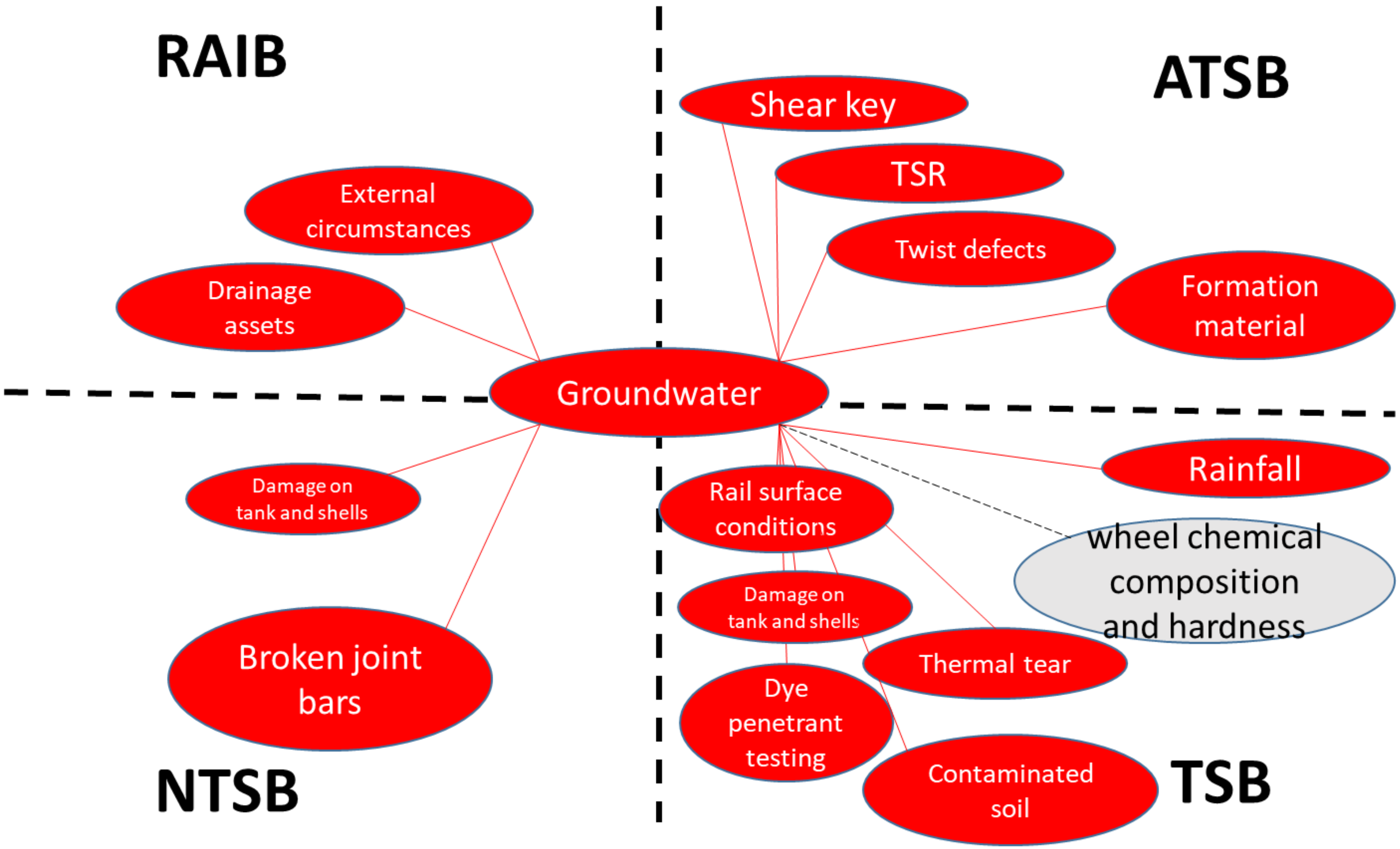


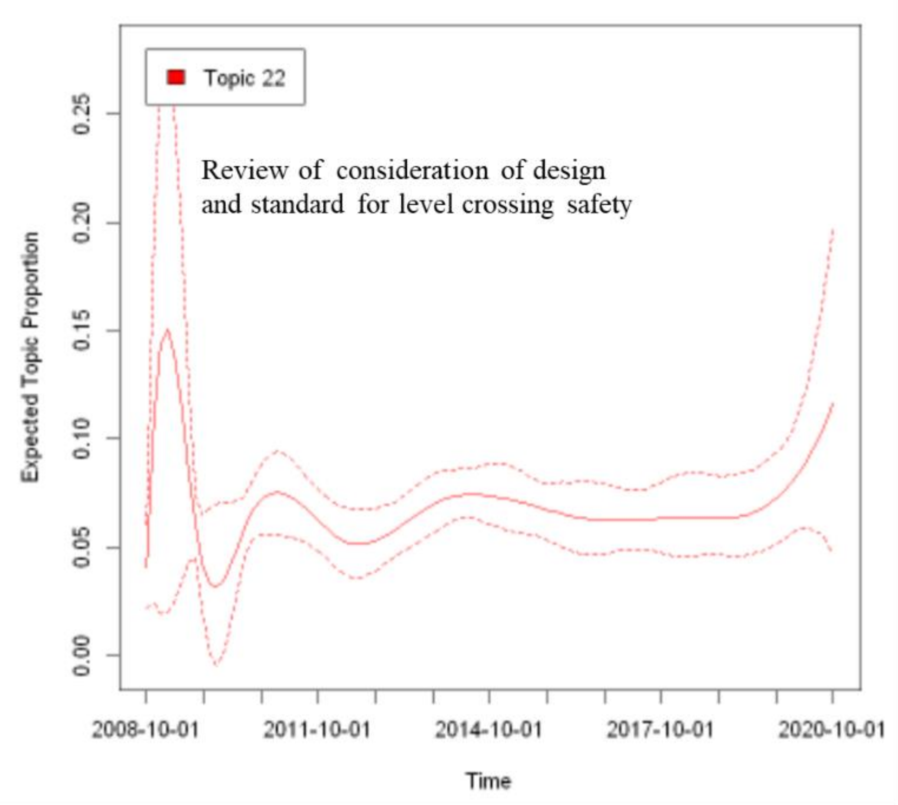
Figure 5-23: An example of a joint hazard "groundwater" combined with other hazards in four countries

5.5.2 Cross-sectional analysis – investigators

The investigating body of each country conducts the railway accident analysis and proposes recommendations in accordance with the nature of the accident to improve railway safety. Learning how recommendations are designed to address hazards in the railway system across countries can be beneficial to prevent similar accidents from occurring in other countries. This section highlights the primary trend of recommendations proposed by each country, and the frequency of co-reference between investigated countries is also discussed. Note that only a significant increasing or decreasing trends are discussed.

Figure 5-24 shows an increasing number of recommendations related to the review process, suggesting more review requirements on existing designs and procedures are found in the RAIB dataset in recent years. Note that the solid line indicates the mean of the probability of the occurrence of the topic, whereas dotted lines refer to the variance. Another noteworthy observation is that several topics relevant to cross-organisational learning behaviour are identified, including process standardisation, communication for lessons learned, and learning from other sources (Figure 5-25). The process standardisation refers to ensuring the consistency of adopted procedures across organisations, whereas communication for lessons learned and learning from other sources mean learning across time and jurisdictions, respectively. An upward trajectory can be found in process standardisation and communication for lessons learned, implying that RAIB highlights the importance of mitigating the difference of applied processes and ensures lessons of previous accidents are addressed to increase railway safety. Recommendations related to learning from other sources were commonly made before 2016 but are rarely suggested after 2016.

On the other hand, a decreasing trend in the number of passive recommendations is found in the ATSB dataset (Figure 5-26), indicating a switch from passive suggestions to distinct instructions made by ATSB to address identified hazards. For the cross-organisational learning-related topic, only the topic of sharing knowledge with organisations is identified. The trend suggests a gradual decline, implying the focus has been shifted to individual control of each railway organisation regarding addressing risks.



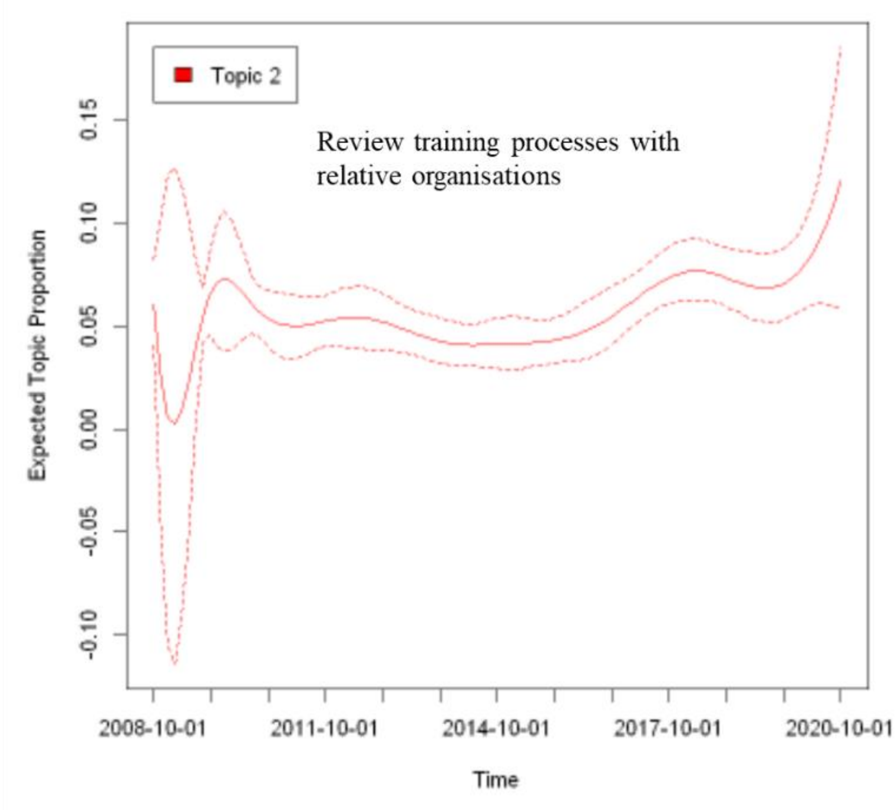


Figure 5-24: The trend of topic 22 and topic 2 in the RAIB recommendations dataset

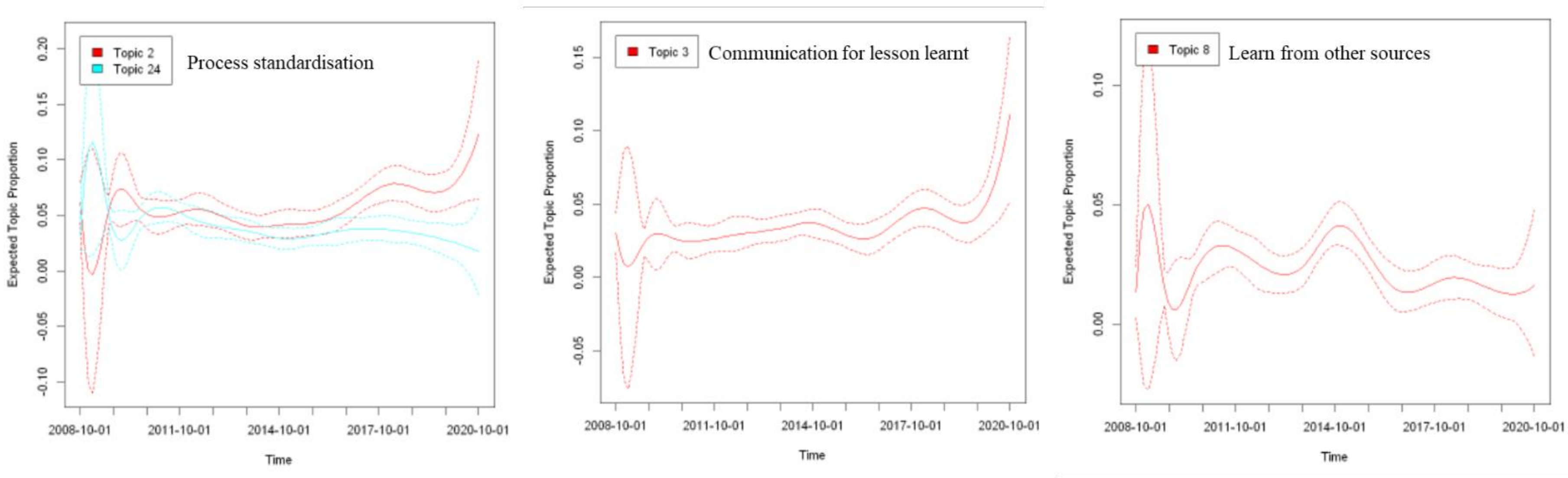


Figure 5-25: The trend of cross-organisational learning topics in the RAIB recommendations dataset

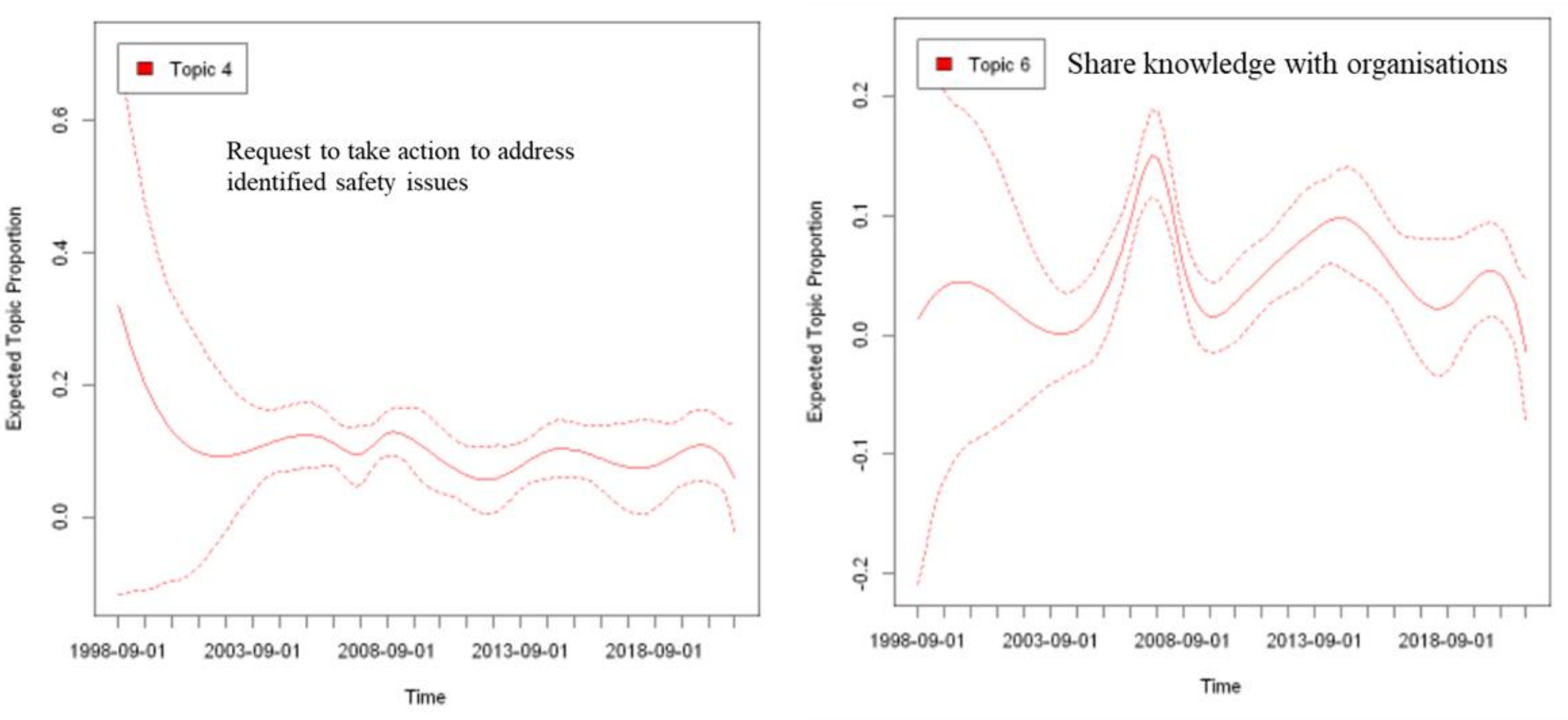


Figure 5-26: The trend of topic 4 in the ATSB recommendations dataset

Next, the result of the NTSB recommendations dataset indicates a sharp rise in the number of recommendations related to the cooperation between local authorities and other organisations (Figure 5-27). There has been no significant trend in making recommendations of introducing new technology or advanced train control systems in cooperation with local authorities before 2000. However, this topic has drawn much attention since early 2017 and maintains a steady increase. On the other hand, a consistent pattern of promoting cooperation with other organisations is observed in topic 5, implying the investigator believes that most identified risks can be addressed by encouraging organisations in the railway industry to share experiences and come up with solutions. For the cross-organisational learning-related topics, three topics are identified and shown in Figure 5-28. A marked fall is identified in assisting research and programs and dissemination (of railway safety knowledge) but cooperating with other organisations is the exception to this trend, showing a slight increase in recent years. This observation might indicate that NTSB believes sharing the knowledge of railway safety and addressing identified risks in cooperation with others can ensure that the railway industry can thoroughly learn lessons. Furthermore, NTSB also devotes efforts to promoting individual research programs' assistance and introduces other resources for different perspectives of addressing railway risks.

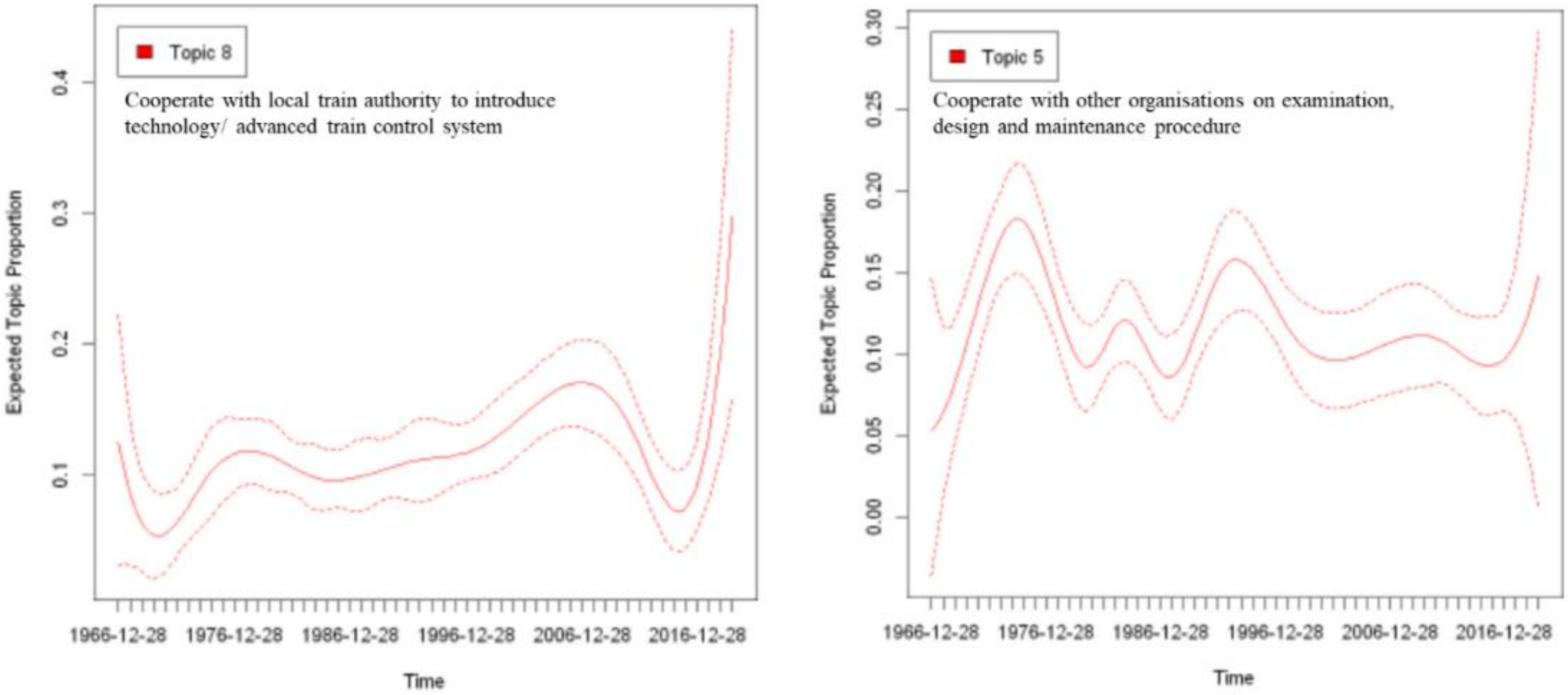


Figure 5-27: The trend of topic 8 and topic 5 in the NTSB recommendations dataset

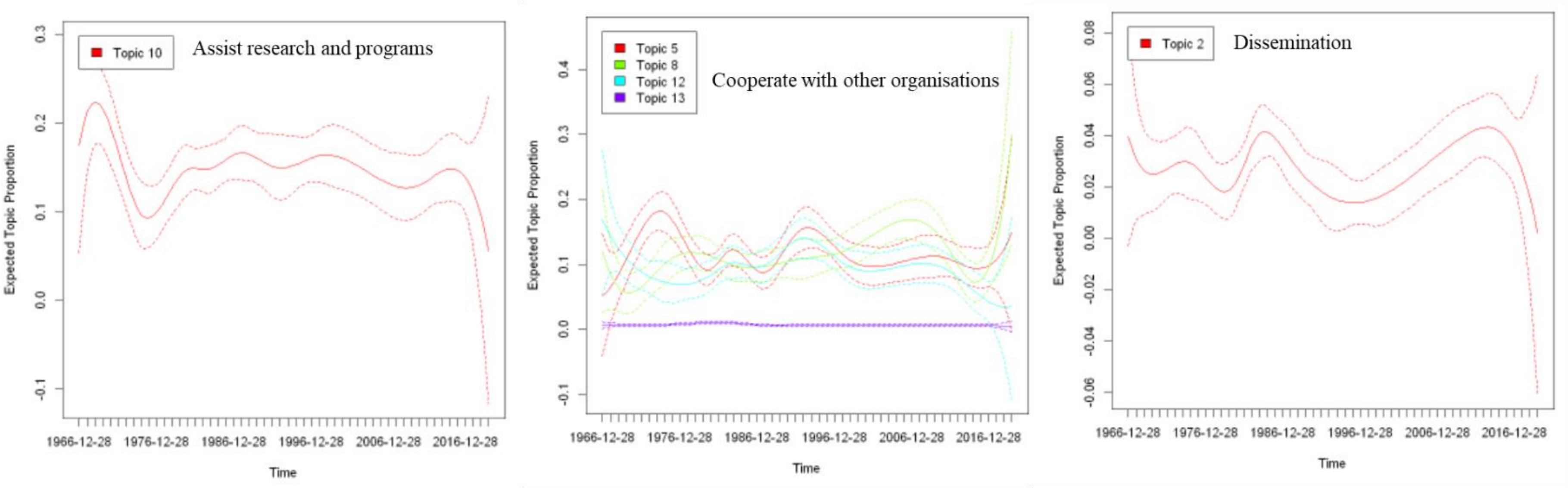


Figure 5-28: The trend of cross-organisational learning topics in the NTSB recommendations dataset

Last, a gradual increase in the number of recommendations related to developing and implementing guidelines and procedures made by TSB is observed (Figure 5-29). The number of these types of recommendations experienced a surge in 2015 and kept climbing after then. On the other hand, a topic related to cross-organisational learning behaviours, “cooperate for standardisations”, is identified despite the relatively low possibility of occurrences. Finally, it is worth noting that the overall variances of the result from the TSB recommendations dataset are comparatively more significant than other countries due to the limited input data.

Overall, the culture of learning across organisations is observed from the investigated countries. RAIB and NTSB put much emphasis on promoting learning across organisations and time given that the increasing number of relevant topics is identified. Although similar topics related to learning across organisations and across time are also found in the ATSB and TSB datasets, a gradual decline in the possibility of occurrences might indicate that less emphasis is placed on this.

Apart from the behaviour of learning across organisations, this study is also interested in the learning behaviour between investigating bodies, which is seldom discussed in the literature. Learning across countries benefits railway safety by understanding unseen risks and finding relevant solutions in advance.

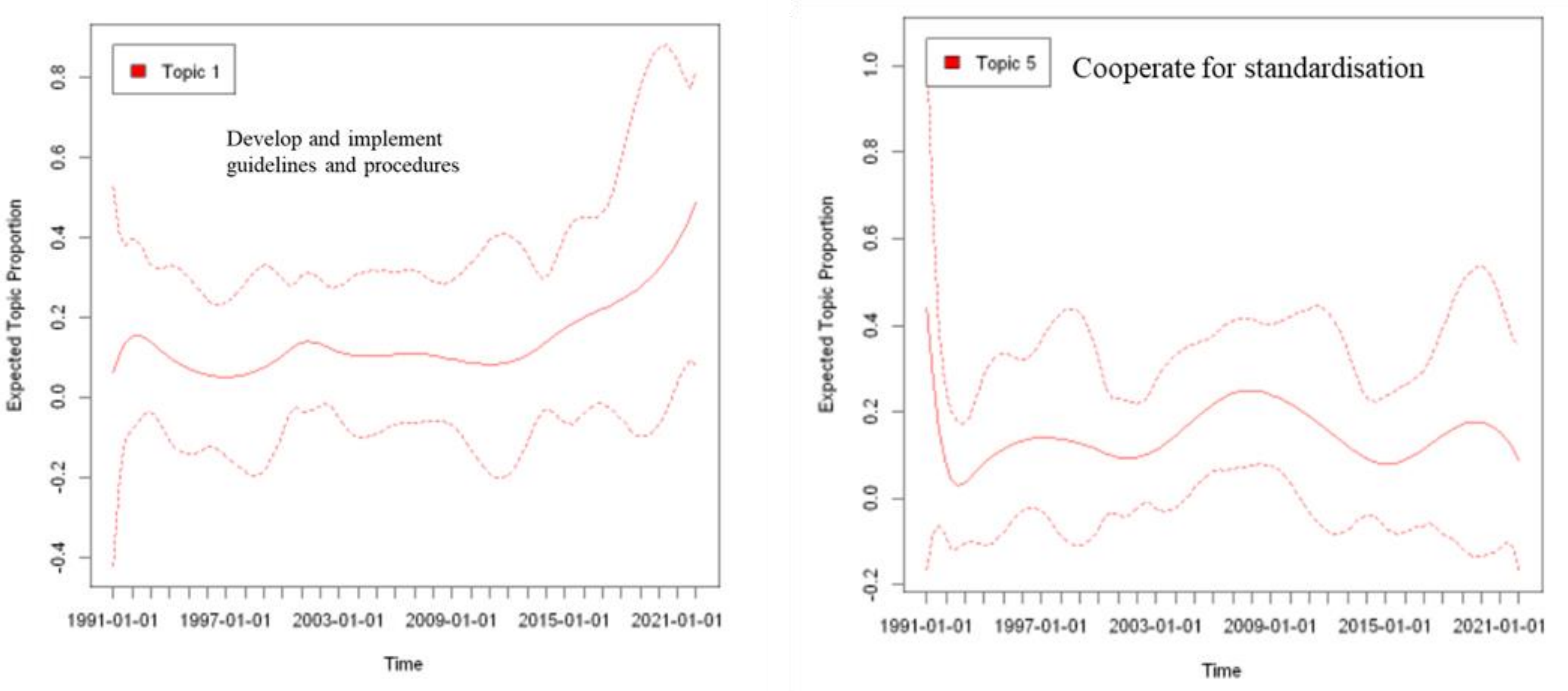


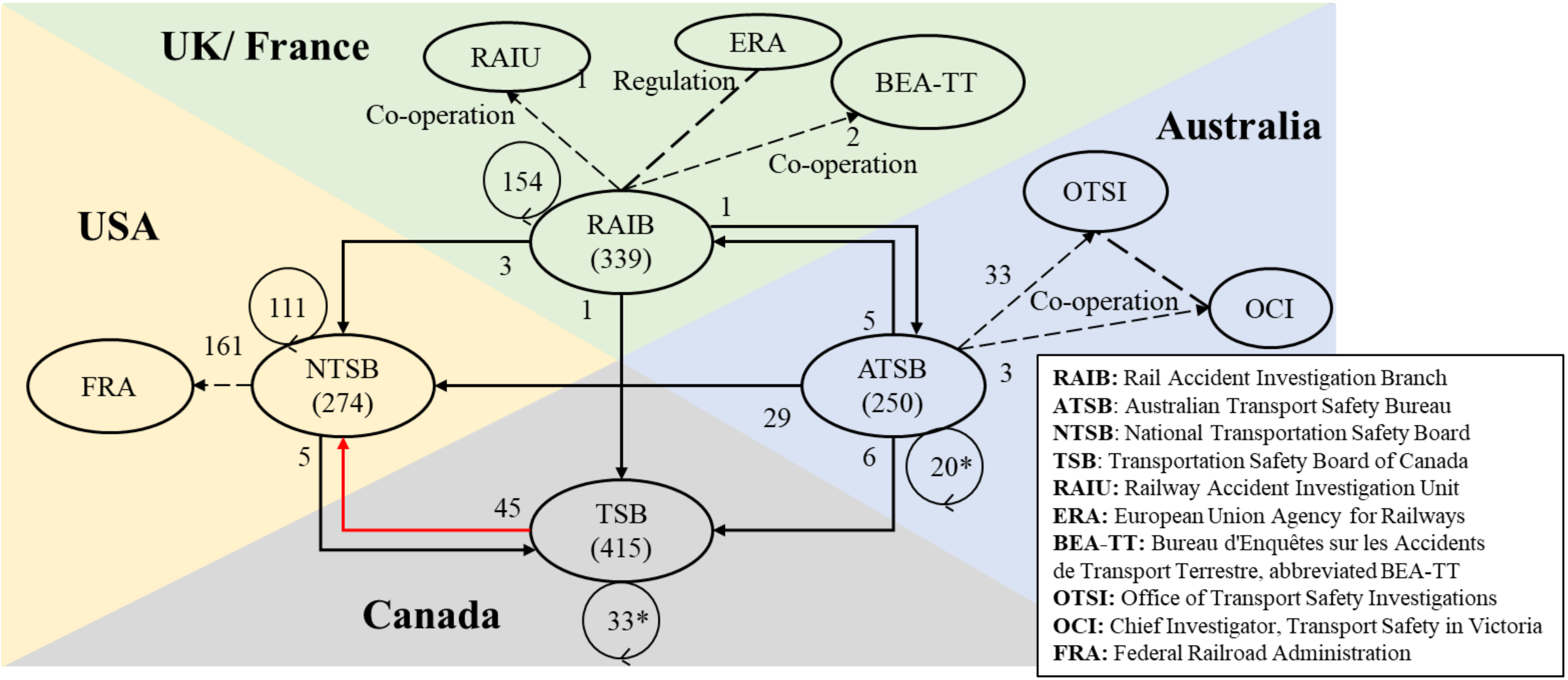
Figure 5-29: The trend of topic 1 and topic 5 in the TSB recommendations dataset

Figure 5-30 illustrates the frequency of co-reference between RAIB, NTSB, TSB and ATSB (solid arrows) and the relation between cooperating organisations within the same area (dotted lines arrows). The number of analysed railway accident reports of each investigator is shown in brackets. The co-reference is identified by counting works published by other organisations in the reference section of a railway accident report. The number under each arrow refers to the frequency of references to the publication of another organisation. The circular arrow indicates the frequency of self-reference.

Of the 415 TSB reports, 45 of them referenced reports published by the NTSB at least once. Of the 250 ATSB reports, 40 reports referenced other investigators. Among all the references, five were to the RAIB, 29 were to the NTSB, and six were to the TSB. Of the 339 RAIB reports, only four reports referenced other investigators, including to the ATSB one time, the NTSB three times and the TSB one time. Last, of the 274 NTSB reports, five reports referenced the TSB.

First, a very limited number of co-references are identified in railway accident reports published by RAIB. However, the number of self-references is the highest compared with others, reflecting the increasing trend of topic 3 “communication for lesson learned” in Figure 5-25. NTSB reports are similar with a high frequency of self-references and low co-references to others. Note that many co-references from NTSB to FRA might not be considered a potential cross-organisational learning behaviour given that NTSB is responsible for reacting to or commenting on actions taken by FRA. Additionally, a comparatively large number of references from ATSB and TSB to NTSB are identified, implying that both investigators consider similar cases that occurred in other countries before conducting the investigation and analysis. Despite the observed behaviour of self-references, the low frequency of co-references might indicate insufficient learning behaviour across jurisdictions.

This section has provided an overview of the recommendations proposed and the trend of each type of recommendation over time across the four investigated countries. A distinct finding is that each country has various cross-organisational learning behaviours, constituting different railway safety learning cultures. The analysis of co-reference and self-reference also shows the same result. Despite the identified different styles of making recommendations, a comprehensive distribution of all topics and the role each topic plays in a socio-technical system are still absent. Therefore, the following sections advance the finding from BERTopic and the STM with the existing theories in the literature to provide another perspective to illustrate the relationship between railway hazards and recommendations made by investigators.



* The occurrences of self-reference might be underestimated because some reports published by TSB and ATSB do not provide reference list or the self-reference is not listed in the reference.

Figure 5-30: The frequency of co-references between RAIB, NTSB, TSB and ATSB (solid arrows) and the relation between cooperating organisations (dotted arrows)

5.6 Evaluation of model – workshops and survey

As discussed in Section 4.2, two scoping workshops and a survey were arranged to ensure that the model developed met the needs of practitioners in the railway industry and to understand potential improvement for further research. The mechanisms and the outcome of the model were presented, and feedback was collected to determine the design of advanced applications. How organisations in the railway industry process the information flow was also observed during the workshops and by the survey.

5.6.1 Outcome of the scoping workshops

The first scoping workshop was conducted with seven participants having a range of experience in the railway industry from 2 years to over 21 years. In the second workshop, six participants had experience ranging from 5 years to over 21 years. The participants represented a diverse range of roles, including consultant, assistant professor, accident investigator, human factor specialist, transport planner, research analyst, and safety programs and initiatives personnel. The number of respondents was limited due to the challenges of recruitment, but the outcome was deemed to be suitable for the purposes of the research. Recognising that this represents only a small group of informed participants, we are satisfied that a suitable range of views prevalent in the industry given the diversity in roles and years of experience was captured within the sample and contributed to a comprehensive understanding of the railway safety industry's challenges and perspectives. Additionally, all participants were fully engaged during the workshops and survey and provided rich and in-depth insights from their practical experience, ensuring robust and reflective findings.

The survey and workshops were primarily conceived as the adjunct to the evaluation of the model developed. However, highly valuable content was observed, such as the potential pattern of learning behaviours of the railway industry and the attitude toward railway safety. For example, costs associated with incidents and the relationship between costs and keywords attracted significant interest during the survey and workshops. Furthermore, participants confirmed that reviewing historical railway accidents would help them understand related situations and the root causes of railway accidents. The selection of reviewed reports is based on the experience and knowledge of investigators. Additionally, several “classic” railway accidents with critical insights or severe consequences were highlighted as frequently cited case studies during the investigation process. These responses might indicate the potential way that the railway industry transfers railway accident reports into knowledge.

For the take-out survey, all participants were encouraged to fill out all questions after the workshops. Six respondents were collected, and all questions were adequately answered. All data collected was associated with the information flow of the railway industry (see Figure 7-1) and used to support the understanding of the learning behaviours in the railway industry.

5.6.2 Analysis and discussion

This section discusses the comments and recommendations collected in the workshops and the learning behaviours observed in the survey. Elements analysed in both sources are compared to ensure the outcomes are interpreted reliably.

5.6.2.1 Workshops

First, several pieces of valuable aspects of feedback related to the interpretation were collected from both workshops while introducing the raw outcomes of the topic modelling. A technical dictionary for the railway terminology and jargon might be required for a better understanding of extracted keywords and topics given that end users might be confused by words not ordered in a logical way. Presenting the outcome systematically or as several summarised sentences might be more desirable. Additionally, the distribution of keywords under a topic presented with the occurrence might be helpful in turning the outcome into human-readable knowledge in the decision-making processes. Connecting keywords extracted to the topic that users are interested in and relative knowledge was also identified as a critical function. Participants made several suggestions regarding the design of diagrams presenting the result of topic models to strengthen the readability.

Practitioners in both workshops also indicated that a number of taxonomy and knowledge systems had been developed and applied in practice. For instance, the UK's Rail Safety and Standards Board (RSSB) has used NLP and machine learning to estimate the risk of railway accidents. A railway hazard list has been developed by RSSB based on their approaches to ensure it is less prone to omissions. The RAIB has also established a memory system storing findings of the previous investigation for statistical purposes, such as the distribution of causes and consequences. On the other hand, practitioners in session 2 workshop pointed out that the Contributing Factors Framework has been published to identify systemic safety issues contributing to an accident and applied to practical railway operation. However, each industry has its own taxonomy of hazards based on the definition of interest, implying that the interface between existing hazard identification and analysis systems in each jurisdiction needs to be addressed before aggregating knowledge from different sources. External conditions that cause the hazard to trigger a railway accident should be considered as well to comprehensively understand the interaction between hazards.

The link between hazards and risk management was also highlighted in the discussion. Participants in the session 1 workshop believed that data has been fully disseminated across jurisdictions because all reports are publicly available. However, converting data into knowledge requires extensive human efforts in consistent analysis and professional knowledge. Specifically, mitigating the gap between systems during the analysis is challenging given that the nature of the system designed and control approaches with various levels of services significantly influence the mechanism and consequences of a railway accident despite the same hazards being involved. Such barriers hinder practitioners from effectively learning from other jurisdictions.

Last, several organisational factors have also been identified as playing a critical role in the knowledge transformation process. The railway industry has been divided into many organisations in various disciplines functioning individually. The focus of each discipline is based on the organisational objective so the interface derived might result in poor cooperation. Such issues might be dynamic due to the change of legislation and regulations, leading to different styles of negotiation between stakeholders during the accident investigation stage and recommendations made. Therefore, the data related to organisational factors collected from historical reports needs to be standardised for further knowledge processing and dissemination.

Overall, participants in the workshops reached a consensus that learning across time has been implemented in the railway industry. Several comprehensive railway taxonomy systems have been developed based on historical railway accidents in each jurisdiction. The limitation lies in heavy human intervention and a labour-intensive analysis process for maintaining and updating existing knowledge in accordance with new railway accidents. On the other hand, practitioners think that learning across jurisdictions might be helpful once barriers to learning are fully addressed. For example, the consistency of data from different sources would determine the quality of the analysis. The standardisation process is also needed to mitigate the existing and external knowledge gap. Organisational factors such as the legislation and regulations over jurisdictions and over time should be considered to understand how they interact with hazards. Given that the complicated nature of a railway accident involves many hazards and underlying factors in a complex system, processing and disseminating such knowledge might become an unaffordable task for an individual organisation.

5.6.2.2 Survey

The take-out survey primarily concentrated on the nature of learning behaviours in the railway industry and how information is transferred to knowledge and disseminated. The questions were divided into four parts: background information, knowledge receiving, knowledge processing and knowledge dissemination (see Section 9.1). Because a minimum 1 year of experience working in the railway industry was a criterion, only eligible workshop participants were invited to fill out the survey. Of the six respondents, three were working in the government sector, two were consultants and one was from academia. Respondents' experience in the railway industry ranged from 6 years to 55 years.

Respondents believe that the "unknown knowns" are the critical elements expected to be retrieved for the information receiving. The source frequently used is primarily jurisdiction-based and known reports or other statistical data within the jurisdiction. The external source is mainly academic papers or analysis reports. Several respondents stated that "we only deal with our own issues/data as it is not transferable to anywhere else" and "Yes, we review accident reports from other countries, review for transferrable learning for [our] Rail, and produce a bulletin summary of incidents and learning", indicating that the transferability of knowledge is the critical element for the railway industry to determine whether to learn across jurisdictions. On the other hand, legislative factors and regulations are another consideration before using knowledge from other jurisdictions because the reliability and accuracy of this information might need to be further validated.

Next, the responses to knowledge processing were diverse based on the role each respondent plays. A UK participant stated that the UK's Rail Safety and Standards Board would help other organisations in the railway industry to analyse railway accident reports and extract valuable knowledge. Another respondent claimed that organisations in the railway industry understand the importance of analysing investigation reports, but aggregated data is unavailable for supporting the decision-making process. Additionally, respondents believed that the time required to review railway accident reports would vary based on professional skills, experience, and the purpose of the review. However, it is agreed that manual review is time-consuming during the analysis. These responses potentially indicate the awareness of this issue but the lack of willingness or ability to process data. In contrast, a consensus view on the expected outcome of processing knowledge is observed, including contributing factors, underlying causes, actions required to mitigate hazards, similar railway accident cases and recommendations made after accidents. To obtain this knowledge, the existing process adopted by the railway industry applies several theoretical supporting tools such as Bowtie, Swiss cheese, Risk control hierarchy and Incident Cause Analysis Method (ICAM).

Last, engagement and information dissemination has been found to be restricted by the legislative framework in the railway industry. Responses like “contractors will often claim legal privilege over a significant investigation making them reluctant to share information relating to that incident”, “difficulties of the willingness to engage (on both sides)” and “difficulties of the access to all appropriate data” indicate that the dissemination and engagement are motivated by contracts or the supervisory relationship within the railway industry. In the UK, the Rail Safety and Standards Board (RSSB) is funded by organisations in the railway industry as the agency body to engage with regulators, such as receiving recommendations, reviewing standards, and conducting safety analysis. Furthermore, the authority-centralised reporting and accident investigating system is observed because practitioners believe that the authority, such as the investigation body and local regulator, would lead the safety improvement process. The legislative framework also regulates the accident disclosure, so the information flow is mostly controlled by norms. Therefore, the process of converting the railway accident into knowledge and disseminating findings is mostly managed by the investigation body at the railway socio-technical system.

5.6.3 Summary of findings from scoping workshops and survey

Material generated by the workshops and survey has shaped the outline of the learning behaviours and the safety culture of the railway industry. However, only concentrating on workshop or survey material might lead to misunderstanding and oversimplifying of the natural behaviour and interaction of a complex system.

The workshops revealed barriers to converting overcrowded external sources into organisational knowledge, such as the lack of proper technical support and legislative framework. The survey further identified the heavy human labour and high professional skills required for knowledge transformation. Furthermore, the restriction placed by the legislative framework of the jurisdiction was frequently mentioned in the survey during the information receiving and disseminating stages. The railway authority primarily determines the use of data sources and information dissemination. Such a supervisory system might limit the motivation of organisations in terms of proactive railway safety improvement.

On the other hand, the workshops and survey also revealed that the information processing system in the railway system is mature in the context of jurisdiction-based analysis. Accident information can be transferred into knowledge by leveraging safety theoretical frameworks and tools and converted into data of interest, such as the causes of accidents and underlying factors. In addition, most information processing tasks have been standardised to ensure the outcome of the investigation is consistent with prior reports.

To sum up, the railway industry has devoted considerable effort to improving railway safety by systematically understanding how a railway accident occurs. Historical railway accidents have been analysed and converted into desirable data forms for further analysis. However, such progress becomes stagnant due to the fast change of external factors and the lack of a modern data analysis process, causing practitioners to ineffectively transform new information into knowledge. Nevertheless, the legislative framework may have resulted in practitioners' concerns regarding handling information flow related to railway accidents and caused the passive reporting attitude. The development of such a safety culture might have a negative impact in the long term.

5.7 Synthesis of findings

This chapter elaborated the application of proposed methods to the four investigated countries. Over 1,200 railway accident reports, containing 400,000 sentences (Table 5-1), published by the national railway accident investigation bodies of four countries were analysed. The indicator coefficient of variance (CV) was proposed and applied to the BERTopic model to discriminate distinguished topics from common ones, enabling the identification of specific hazards. Scoping workshops and a survey were conducted to collect suggestions from practitioners in the railway industry to understand further improvement on the model developed. Various types of hazards were identified by each country, showing the difference in encountered mechanisms of railway accidents and the approaches to address them. For instance, tram-related accidents dominate the RAIB dataset, whereas freight train-related accidents are more common in the ATSB dataset.

Additionally, the STM was applied to the recommendation dataset of the four countries. The trend of making recommendations and the focus of each investigating body were revealed. The RAIB tends to make recommendations relevant to improving existing systems by reviewing processes, regulations and standards, and updating the monitoring or measuring approaches. On the other hand, the ATSB proposes a significant number of recommendations requesting to take action on safety issues identified without direct instructions, offering high flexibility to the railway industry in decision-making. Next, recommendations about cooperating with other organisations and local authorities are more common in the NTSB dataset, suggesting a solid promotion of learning and sharing knowledge between organisations in the railway industry in the US. Last, Canada's TSB focuses on the measurement of ensuring effective procedures and designs are in place to mitigate hazards.

Overall, recommendations related to the introduction or modification of procedures and monitoring systems without discussing the real motivation of behaviours, indicating systematic reviews of the railway safety culture are absent. In addition, reviewing compliance with existing procedures or regulations implicitly indicates a blame culture that is the opposite of the promises made by investigators. Nevertheless, the observations of effectiveness and the absence of day-to-day analysis, operation and maintenance are frequently proposed without discussing the underlying factors or synthesising previous accidents and findings. The effect of such recommendations might not be beneficial for long-term railway safety due to the absence of inherent motivation.

To conduct the cross-sectional analysis, the ontology, knowledge graphs and entity linking were integrated into the designed railway accident ontology to standardise the different terminology used in the investigated countries. Mentions of one entity can be identified by leveraging open access API and SPARQL query service without constructing a complicated named entity recognition model. The knowledge structure Wikidata was applied to mitigate the interface barrier between railway accident reports and the railway accident ontology, enabling us to extract critical elements in railway accident reports by using the context-sensitive disambiguation process.

Participants in the scoping workshops and survey revealed several issues. The major concerns discussed were the interpretability of the model's outcome and transformability of data from external sources. The next chapter developed conceptual frameworks to address issues mentioned.

6. Novel conceptual frameworks proposed

To comprehensively understand the relations between hazards, accidents and recommendations, this study further extends the findings retrieved from BERTopic and the STM by incorporating knowledge from other literature. A holistic view of the subject matter enables the identification of mechanisms and patterns that might not be revealed with a single data source. The results of advanced applications can be beneficial for developing machine learning based interpretations and improving railway safety. In addition, comments and suggestions received from the scoping workshop are also considered during the development of advanced applications.

6.1 The relations between hazards, accidents and recommendations

There have been a considerable number of studies discussing the relationship between hazards and accidents in the literature, such as the domino theory (Heinrich, 1941), Reason's Swiss cheese model (Kim & Yoon, 2013; Zhou & Lei, 2018), and systems theory (Systems Theory). In recent years, growing attention has been drawn to interoperating the existing findings of railway accident mechanisms from a systematic perspective. For instance, the widely used systems theoretic accident model and process (STAMP) assumes that accidents are dynamic and complicated processes and direct and indirect control and causality are considered. Accidents are caused by inappropriate management, and the involved socio-technical system is a control system, offering safety constraints to manage hazards (Gong & Li, 2018; Hulme et al., 2019; Ouyang et al., 2010). Another example is the *AcciMap* based on the system-theoretic process analysis (STPA), the causal analysis based on STAMP (CAST) and the Reason model. The *AcciMap* provides a holistic approach to showing interrelationships between causal (hazardous) factors at different levels and highlighting the problem areas. An exhaustive review of these foundational theories and risk analysis frameworks in the context of risk analysis context was provided in Section 2.2.

However, the discussed relationship does not involve the solutions applied to address risks. Many recommendations proposed by investigators aim to address hazards and prevent similar accidents from occurring, but only a limited number of studies put emphasis on building a comprehensive view of applied solutions worldwide, hindering practitioners and researchers from learning across jurisdictions and across time. Therefore, this study leverages the system theory and the outcome of our models to develop a novel framework for understanding the relationship between hazards, accidents and recommendations and illustrates the process in the following sections.

6.2 RecoMap – a systematic view of recommendations

The argument about modelling risk management by considering it to be a control problem and addressing issues from the perspective of a control structure inclusive of the society for each type of hazard begins with the work done by Rasmussen (1997) and has been ongoing for decades. Several studies have applied such a concept to real-world cases and proposed frameworks for interpretation (Arenius & Sträter, 2014; Grant et al., 2016; Yuyua et al., 2021). *AcciMap* is one of the commonly used frameworks representing the interactions between hazardous elements from different systems in a structured way, such as technology, human factors and environment (Stanton et al., 2019; Thatcher et al., 2019; Underwood & Waterson, 2013; Wheway & Jun, 2021).

Figure 6-1 illustrates the structures of *AcciMap* and the included elements, identifying the potential improvement rather than the responsibility for the railway accident. *AcciMap* also elaborates on the decision flow, including consequences and reactions from the top to the bottom of the system. However, this framework might not be applicable to the railway recommendations data because the shape of one recommendation covers several system levels. For example, a recommendation for a railway accident suggesting learning across organisations might involve system levels at regulatory bodies, local governments and the railway industry. It might be difficult to implement this recommendation in the existing *AcciMap*. Nevertheless, *AcciMap* cannot visualise multiple types of recommendations and cannot describe the trend of decisions made by one organisation over time, hindering users from having a holistic map of all recommendations made by different countries and making the comparison.

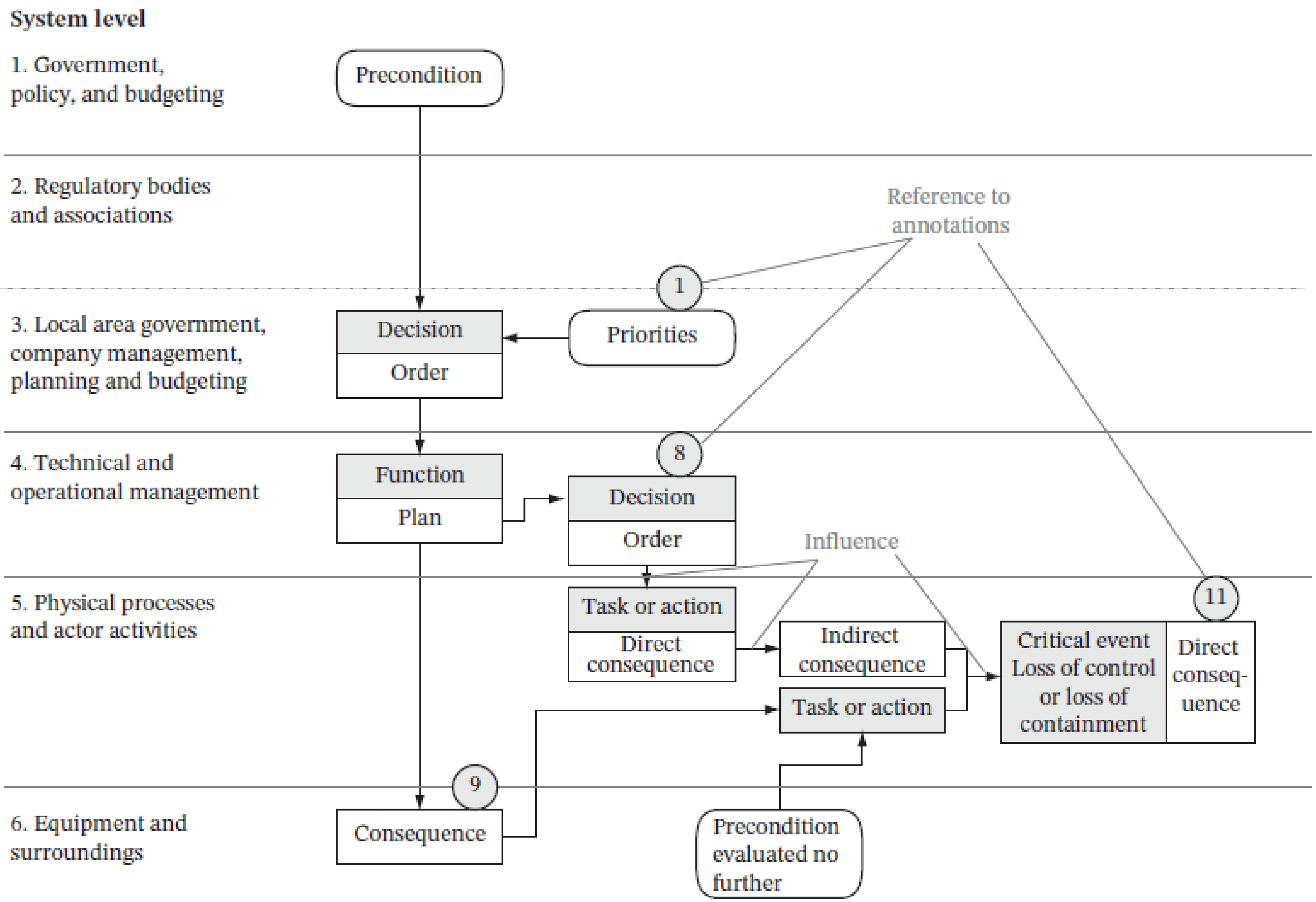


Figure 6-1: The structures of AcciMap and the included elements (Rausand, 2013)

To better understand the style of each investigator’s recommendations, the taxonomy of recommendations proposed by Karanikas (2016) is used to discriminate the recommendation type based on the extent to which the railway has the flexibility to address hazards identified by investigators. There are three types of recommendation proposed by Karanikas: assignment, action and reminder (Table 6-1). Assignment offers a distinct objective for organisations to come up with solutions and implementations and is considered a supportive recommendation. In contrast, an action might contain specific approaches assigned by the investigator to address hazards, limiting the flexibility of organisations to adopt solutions and be categorised as an interfering type of recommendations. Last, the reminder is another supportive recommendation, providing enormous flexibility to organisations in modifying the existing rules and procedures of the operation.

There is no one-size-fits-all solution for addressing hazards in the railway industry given that each country uses a wide range of systems and has developed an inherent railway safety culture. In addition, recommendations also need to reflect the nature of the investigated railway accident and should be balanced between each type to ensure moderate flexibility in implementing solutions.

Table 6-1: Types of recommendations made (based on Karanikas, 2016)

Recommendation type	Description	Example	Role
Assignment	Assign an objective for organisations to resolve identified hazards	Network Rail should identify and implement suitable measures to mitigate the risk of a runaway.	Supportive
Action	Assign specific methods to address identified hazards	Network Rail should amend its National Hazard Directory to include the access point alongside South Hampstead station.	Interfering
Reminder	Remind about compliance with existing rules or procedures	FRA should increase monitoring of their employees for compliance with existing applicable rules and procedures.	Supportive

Next, this study slightly modifies the method of describing the role each recommendation plays in the railway system and proposes the customised model *RecoMap* to address the mentioned issues. Instead of showing the decision flow, *RecoMap* enables a variety of recommendations to fit at multiple system levels and describes the trend of each type of recommendation across different countries. Figure 6-2 shows the proposed *RecoMap* applied to the outcomes of the STM. The extracted topics are placed in the *RecoMap* in accordance with the covered systems, and the number of occurrences is labelled as well. The depth of the colour represents the time that most recommendations are sorted to the topic proposed. The light colour refers to early recommendations and dark colour to late recommendations. The railway system is divided into the organisational level and operational level, representing how the socio-technical system works in the railway industry. Therefore, *RecoMap* addresses the concern about aggregating findings and insights obtained from railway accidents in other jurisdictions, which was proposed by practitioners in the scoping workshop. Furthermore, practitioners can review what role each recommendation plays in the socio-technical framework and understand how the legislative framework and regulations influence railway safety in each jurisdiction by implementing the recommendations made.

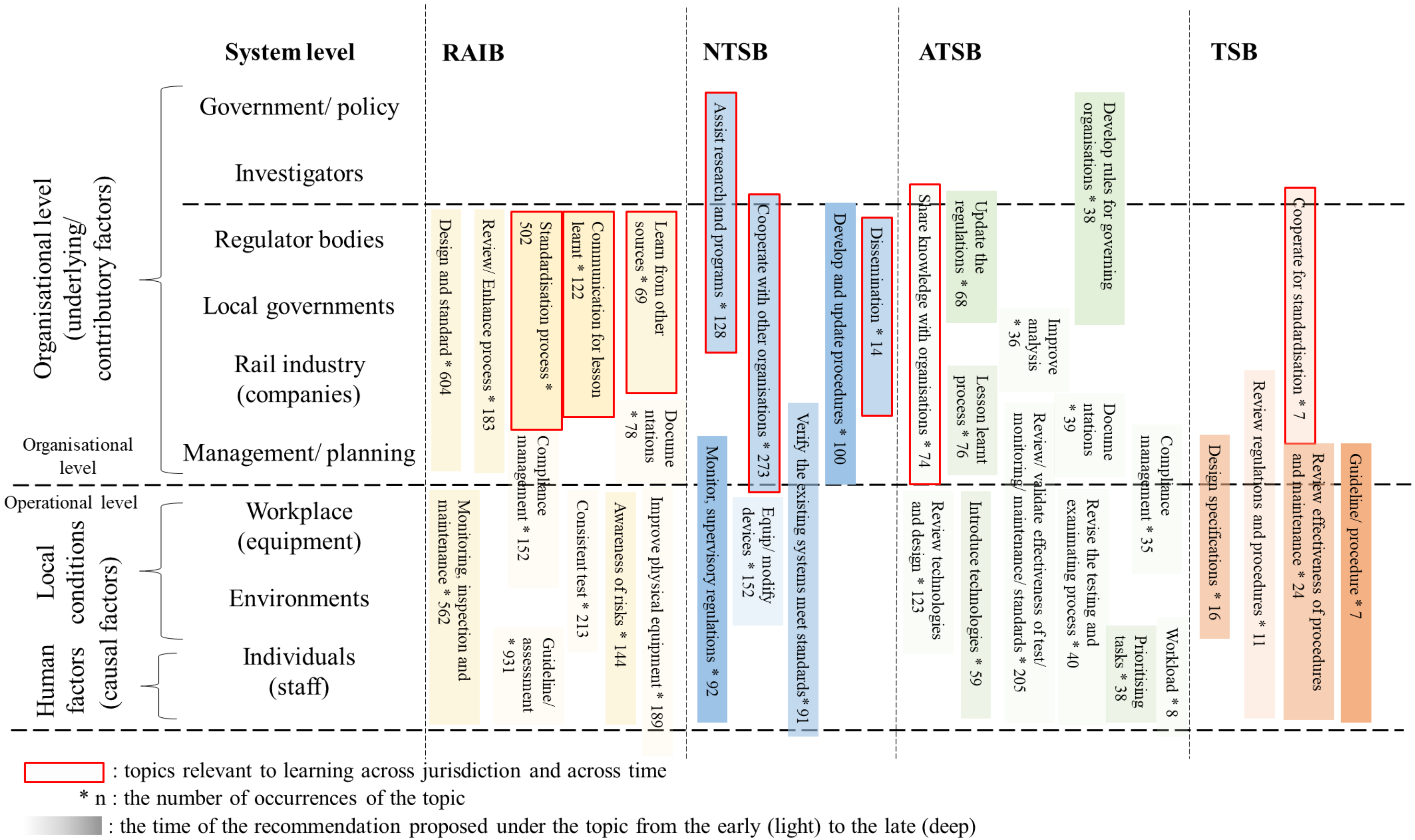


Figure 6-2: The proposed RecoMap applied to the outcomes of the STM

Overall, *RecoMap* maps how investigators in different countries address identified hazards and lead the railway industry of each jurisdiction to improve railway safety. Common recommendations at the operational level are procedures of maintenance and inspection, consistency of testing processes, introducing state-of-the-art equipment, and reviewing existing designs and technologies. On the other hand, recommendations commonly proposed at the organisational level are process standardisation, cooperation with other organisations and dissemination of railway safety knowledge.

A growing shift from addressing hazards at the operational level to the organisational level is found in ATSB and RAIB recommendations, implying the railway industry is gradually adopting system theory and control theory to improve railway safety and address risks from the perspective of an integrated whole. Such a trend might be beneficial in changing conditions for a system as complex as the railway system, providing useful predictive capabilities to adapt the railway system to a dynamic environment. On the other hand, NTSB consistently offers recommendations at the organisational level, in contrast to recommendations made by TSB.

Several recommendations made by ATSB indicate detailed instructions at the operational level, such as the prioritisation of tasks, the management of workload and validation of the effectiveness of existing standards. This might imply that ATSB tends to propose interfering recommendations (the action type in Table 6-1) to address identified issues. However, most interfering recommendations are coloured in light, indicating a potential shift toward supportive recommendations at the organisational level. A similar shift can also be observed in RAIB, transferring from interfering recommendations such as improving physical equipment and assessments of individuals to design and standardisation of the system.

On the other hand, a considerable number of recommendations related to the cooperation between organisations and assisting research and programs are proposed by NTSB, implying a strong promotion of learning across jurisdictions and sharing the knowledge with other research organisations. The trend continues, along with recommendations relating to the dissemination of railway safety knowledge. It is also worth noting that NTSB consistently tends to propose precise but interfering recommendations, such as verifying existing systems and assisting research and programs.

Last, the number of recommendations made by TSB is extremely limited because only their investigation of major railway accidents results in recommendations. In addition, most recommendations address hazards from the operational perspective and rarely are recommendations at the organisational level proposed, hindering how the railway industry deals with hazards as an integrated whole.

Different combinations of the style and system level of recommendations might be feasible for different roles in the railway industry. Therefore, the roles that local railway regulators and national railway accident investigators play should be differentiated. For local regulators, operational recommendations might be appropriate to be proposed given the high homogeneity of railway systems and operations. In addition, local regulators have more experience and understanding of railway systems in their jurisdictions. The level of cooperation is higher than national railway accident investigators, indicating that interfering recommendations might be more efficient to improve railway safety.

On the other hand, national railway accident investigators are able to instruct the whole railway industry, including local railway regulators. Therefore, the emphasis should be put on proposing a positive railway safety culture, disseminating railway safety knowledge and ensuring lessons are fully learned and applied to all relevant railway organisations across the country. In doing so, the recommendations made by national railway accident investigators need to be supportive, offering organisations the best flexibility for local railway regulators and railway organisations to modify the day-to-day operation and gradually adopt new approaches to manage potential impacts. Furthermore, recommendations at the organisational level are suggested to be proposed by national railway accident investigators to enhance the communication and safety culture of the whole railway industry. Promoting learning behaviours and reducing the obstacle of the interface between organisations are also critical objectives to be achieved. Thus, proposing supportive recommendations to address hazards and manage risks from the organisational perspective might be most beneficial for railway safety from a national railway accident investigator.

Table 6-2 shows the comparison matrix for the investigated countries between the style and system level of recommendations. Cells from the top left (dark grey) to the bottom right (light grey) represent the suggested combination of the style and system level of recommendations adopted by the investigator at the lower system level (i.e., local railway regulators) to the higher system level (i.e., national railway accident investigators). Each investigated country has been divided into two stages: the early stage and the current stage. Overall, the style of proposing recommendations of all countries at the early stage tends to be interfering at the operational level. Such a trend has gradually shifted to making supportive recommendations at the organisational level. However, the majority of recommendations made by NTSB are still interfering, and TSB thus far proposes most recommendations at the operational level. Therefore, it is suggested that investigators at different levels consider the role they play in the railway industry in their country before developing recommendations.

Table 6-2: Comparison matrix for investigated countries between the style and system level of recommendations

Style / System level	Interfering	Neutral	Supportive
Operational	ATSB (early years)	RAIB (early years) TSB (early years)	TSB (current)
Neutral	NTSB (early years)	RAIB (current)	ATSB (current)
Organisational	NTSB (current)		

Apart from the style of making recommendations, learning behaviour also plays an essential role in advancing railway safety (Paul et al., 2018; Zhan & Zheng, 2016). Topics related to the learning across jurisdictions and across time are highlighted with red outlines (Figure 6-2), including lesson learned, communication, dissemination and cooperation. The result suggests that investigators gradually put emphasis on exchanging knowledge and learning across organisations within the jurisdiction in recent railway accident reports, indicating that the adjustment to correct mistakes (single loop learning[6]) and the identification of underlying factors (double loop learning[6] have been fully implemented in the railway industry. However, the participation of people in making well-informed decisions to address complicated and dynamic risks (triple loop learning[6]) is not yet found in made recommendations. For instance, recommendations seldom review cultural dimensions. The idea of learning across organisations has been proposed, but investigators rarely remind the railway industry to

[6] The concept of single loop learning, double loop learning and triple loop learning is proposed by Romme and VanWitteloostuijn (1999), suggesting that traditional organisational learning and design (single and double loop learning) focus on risks having simple structural patterns. To understand and address the complexity and dynamics of changes, triple loop learning should be adopted to explore structural opportunities and promote the participation of people in making well-informed decisions.

understand the value of making these decisions and might result in a passive attitude toward railway safety.

To sum up, the proposed *RecoMap* provides a holistic view of how different countries make recommendations, enabling the railway industry in other countries to learn potential approaches to address similar risks from these countries. The style of making recommendations suggests that the most appropriate style for each organisation might vary based on its role in the railway system. A shift from making interfering recommendations at the operational level to making supportive recommendations at the organisational level was also identified among investigated recommendations made by national railway accident investigators. Last, the learning behaviours were also observed, and the analysis suggests that the behaviour of triple loop learning is still insufficient in the railway industry of the investigated countries. Learning from recommendations might not be the only way to improve railway safety but understanding recommendations can help the railway industry understand how similar issues are addressed in other jurisdictions.

6.3 The development of HazardMap

Despite distinct topics identified by the BERTopic model, the relationship between each topic cannot be revealed, hindering users from understanding the mechanisms of railway accidents. Nevertheless, a considerable number of comments collected from the scoping workshops indicate that the outcome is difficult to interpret by merely reviewing the keywords of each topic. The discussion also covered that the nature of railway accidents is characterised by a series of hazards, so it might be unrealistic to look at an accident through a singular perspective. Therefore, the result of the BERTopic model is further extended by adding additional processes to address this issue. First, the distribution of the number of sentences over each topic on documents is extracted and condensed to a topic-document matrix. Second, we assume that the distribution of each topic over documents is the projection of the extent to which this topic influences each railway accident. Multiple similar distributions indicate that these topics constitute a specific group of railway accidents with similar features. Therefore, the cosine similarity approach is applied to identify the similarity of distributions (Cheng et al., 2009; Qurashi et al., 2020). A topic–topic similarity matrix can be generated with each element between 0 and 1. A higher similarity score indicates that sentences under both topics are commonly used in the same group of documents.

Next, a distribution of topics including the relationship can be mapped by setting a threshold for the similarity score, linking each topic and forming a series of clusters representing various hazards. The threshold for the similarity might be determined based on the nature of input data and analysis purposes. A higher threshold leads to scarce links between topics and forms a limited number of small clusters, whereas a lower threshold results in dense connections between topics and several large clusters containing almost all topics. Therefore, the threshold needs to be carefully determined by reviewing each outcome with a different similarity score.

This study uses the RAIB dataset as an example for demonstrating the application. The threshold for the RAIB dataset is set as 0.5 due to the appropriate balance between the number of clustered groups and the well-distributed hazards. Figure 6-3 shows the distribution of relationship between hazards identified in the RAIB dataset. Each orange dot represents a topic identified by BERTopic and the link refers to the similarity score of two topics that is larger than the threshold. The name assigned to each cluster is based on the inference of keywords of linked topics. According to this result, more potential hazards are identified compared with the interpretation of topics having high possibilities of occurrences. The connection between topics is also revealed to illustrate the underlying causal relations in the hazard group. In addition, the cross-country analysis becomes applicable by extracting the hazard of interests from different countries and comparing the mechanisms and causal relations. Two case studies are illustrated in the following sections.

On the other hand, similar accidents still occur given that recommendations are made and adopted by the railway industry. Such situation might not be entirely relevant to the issue of how hazards are addressed but the way hazards are interpreted, indicating the need for a re-interpretation of hazards in the railway safety context. Additionally, the distribution of the relationship between hazards constitutes each cluster by aggregating connected hazards, indicating the nature of the complexity of a hazard. Therefore, the hazard should be interpreted by elements involved rather than the hazard itself given that it would trigger another accident in combination with other hazards or in other dimensions.

Assume that each hazard in the railway system is revealed in the form of accidents. A hazard must have multiple aspects that result in different types of accidents. Therefore, an addressed hazard might appear again after the combination with others, implying one hazard will never be fully addressed. However, reducing the caused impact by proposing appropriate recommendations toward accidents revealing part of the aspects of one hazard is still beneficial for improving railway safety.

Based on this description, we can assume that each hazard has almost infinite aspects. For example, one aspect might result in an accident with the combination of other hazards and under specific conditions. Once the accident occurred, the impact would disrupt the railway system and recommendations are proposed to address the triggered aspect of this hazard by the investigator. After several occurrences of accidents and all controllable aspects are addressed, this hazard is considered to be mitigated to the lowest possible level.

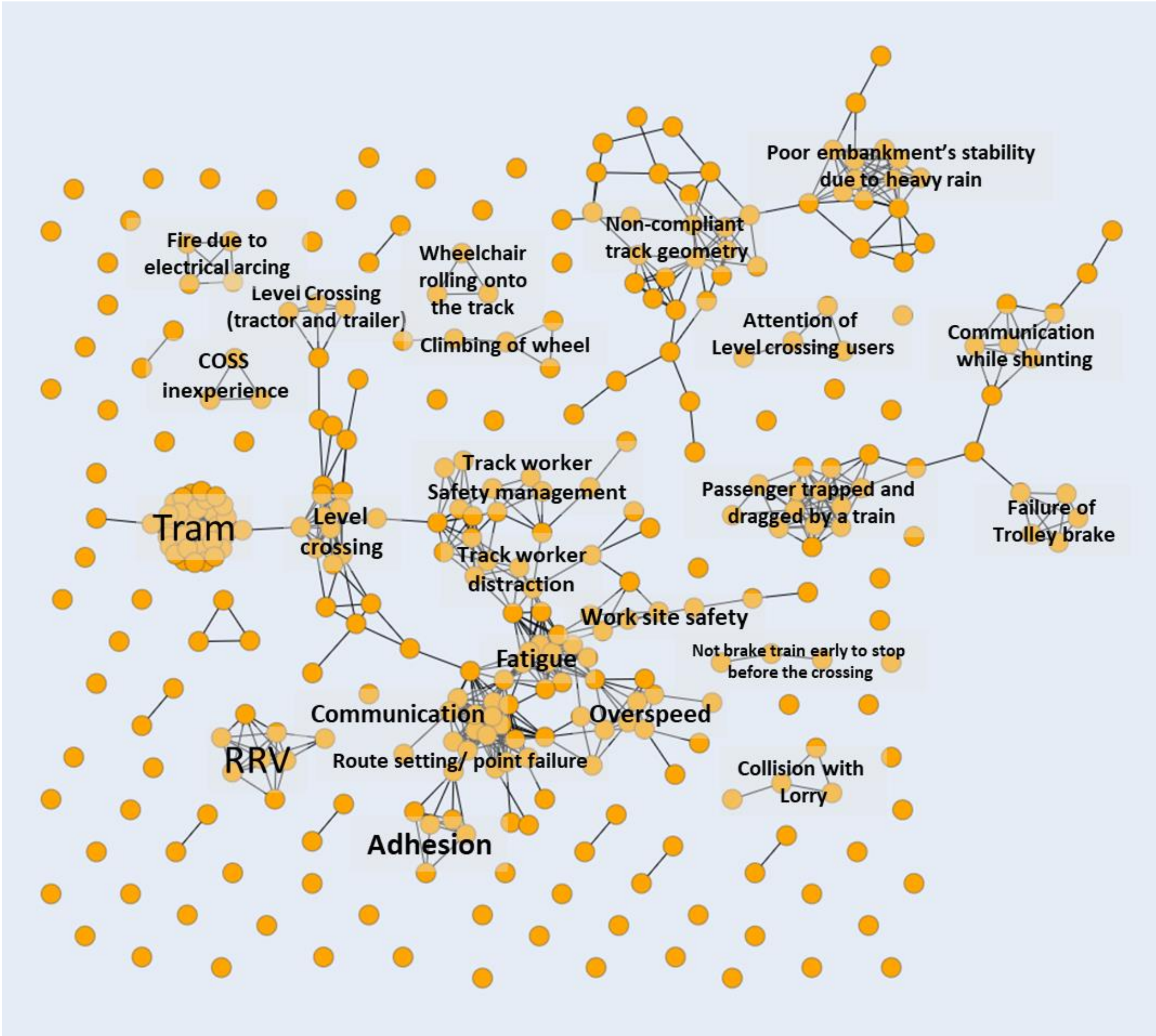


Figure 6-3: Distribution of the relationship between hazards identified in the RAIB dataset

To visualise the description above, this study proposes *HazardMap* to help better understand the relationship between hazards, accidents and recommendations across countries. Figure 6-4 illustrates the conceptualised framework *HazardMap*, inspired by the result of topic modelling (Figure 6-3). *HazardMap* is a data-driven and epidemiological factor-based framework, looking at railway accidents from a hazard-centred perspective. Hazards illustrated in *HazardMap* are derived from clusters of hazards, for example, the level crossing in Figure 6-3.

Next, each hazard has a series of aspects illustrated as the outline of the oval consisting of continuous dots. Two types of dots surround the hazard: the unprotected aspect (coloured in red) and the protected aspect (coloured in black). The unprotected aspect refers to the potential possibility that this hazard triggers a railway accident under specific conditions or in combination with other hazards. The unprotected aspect might not be identified until it triggers an accident or preventative implementation is placed in advance. On the other hand, the protected aspect represents the hazard that would no longer trigger an accident from this dimension because it has been identified and fully addressed by introducing permanent solutions, such as applying state-of-the-art technology or improving relevant processes. Note that any implementation of new policies or strategies might result in another hazard while fully addressing an aspect of a hazard.

Once a hazard triggers a railway accident (with the combination of other hazards or factors), the aspect would be highlighted in the *HazardMap* and connected to the triggered railway accident. Multiple aspects of hazards might trigger some railway accidents, for instance, accident 11 is triggered by aspects of hazard 1 and hazard 2. Subsequently, railway accident investigators would investigate and propose recommendations to address identified aspects of the hazard involved, aiming to prevent similar railway accidents from occurring again (converting red aspects into black). Some recommendations might also address hazards identified by previous railway accidents and reinforce the prevention of hazards, which is illustrated as multiple arrows toward different accidents in the *HazardMap*.

On the other hand, hazards are further categorised based on likelihood across countries. Some hazards can only be found in specific locations, such as autumn leaf falls in the UK and high temperatures hazards in Australia (see Table 5-8). On the other hand, hazards that can be classified into more than one country are shown as common hazards. Different aspects of these common hazards might affect the country-specific area and trigger a railway accident. Note that the locations of hazards on the *HazardMap* might move from one area to another to reflect the change in environment. For example, due to severe climate change the high temperature hazard might impact the UK railway system. In this case, the high temperature hazard might move from Australia to the common hazard area.

Last, *HazardMap,* with the use of the ontology developed, also addresses several issues raised in the workshops and offers an effective approach to incorporate data retrieved from external sources. First, the technical dictionary for the railway terminology and jargon has been covered by the ontology. Although the model cannot directly offer a full list of terminology and jargon used in the dataset, the ontology enables users to conduct analysis across jurisdictions by standardising the terminology used in the topic of interest. Second, the existing hazards taxonomy developed by each railway industry can be further connected to the *HazardMap* across countries to discover potential aspects of hazards that are overlooked in the system. Finally, *HazardMap* also provides a consistent and standardised framework for practitioners to process and archive the railway accident knowledge acquired from the investigation with limited human intervention required.

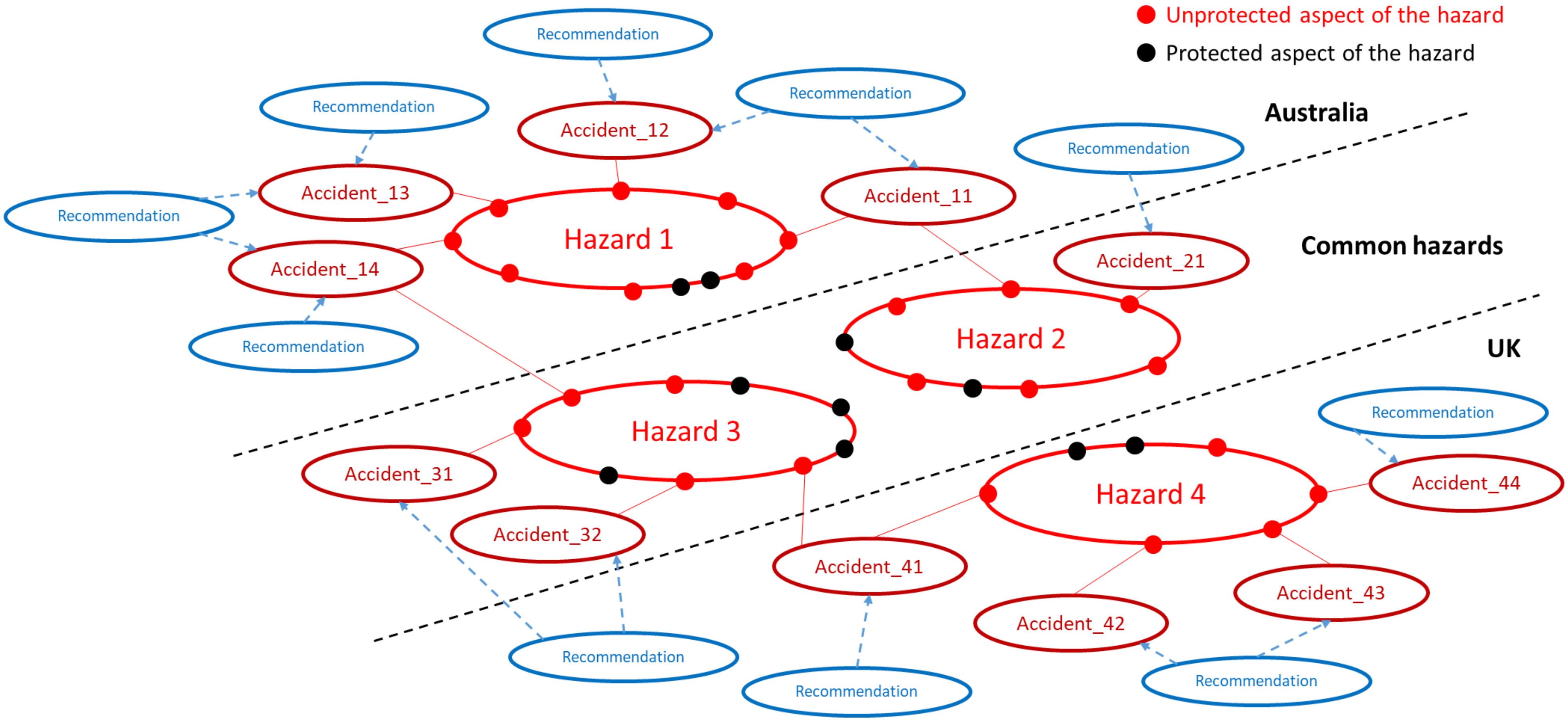


Figure 6-4: HazardMap: The relations between hazards, accidents and recommendations across two countries

6.4 Temporal analysis – the time required for making recommendations

Another potential hazard, which cannot be adapted to the *HazardMap* and has been overlooked in existing recommendations, is the gap between the occurrence of the accident and the publication of the recommendations. Aspects of hazards triggering one accident are identified and addressed after recommendations are made and implemented, but the railway system is exposed to these hazards during this investigation and reporting period. Although the time required to make recommendations depends on the nature of the investigated railway accident and immediate improvements might be in place as reactions, hazards at the system level still need comprehensive solutions and holistic strategies to be addressed. Therefore, railway accident investigators can either reduce the time required to make recommendations or consider recommendations made by other countries before the accident occurs.

To understand the distribution of the time required to make recommendations and the method adopted to address hazards, this study further extends the *RecoMap* by combining types of recommendations and the time required to publish the railway accident report. Figure 6-5 uses level crossing incidents investigated by the RAIB as an example to visualise how recommendations made in railway accidents relevant to level crossings and the different roles each recommendation plays in the *RecoMap* are distributed over time. According to Figure 6-5, railway accidents relevant to level crossings require 1 to 2 years for investigation, and most recommendations concentrate on the design and standards for level crossing safety (topic 22). The promotion of appropriate measures for monitoring (topic 5) can be found between 2009 and 2015, implying that monitoring systems such as CCTV might be in place during this period and address some aspects of level crossing hazards. This reduces the occurrence of level crossing accidents from such aspects, and the same pattern can be found in recommendations related to adding considerations to existing guidelines and assessment (topic 13).

Note that the time of publication is used as the proxy of the time that the local railway industry receives recommendations given that the real informing time is barely available and each country might have various regulations on the interaction with other organisations during the investigation. In addition, Figure 6-5 only illustrates level crossing accidents investigated by the RAIB and cannot represent all level crossing accidents in the UK because the national railway accident investigation body investigates only a limited number of them.

In summary, hazards in the railway system emerge along with many factors, such as the design of infrastructure, the safety culture, the way of operation and planning. Railway systems in different countries generate inherent hazards and aspects. However, similar hazards can still be found across countries, such as human error and track defects. Many of these hazards and aspects have been thoroughly discussed and mitigated in well-developed railway systems, and the experience can be used to mitigate the hazard before the occurrence of accidents. Therefore, *HazardMap* is proposed to describe the nature of railway accidents, allowing end users to have a comprehensive view of all identified aspects of hazards and corresponding recommendations around the world. In addition, the hazard of the gap between the occurrence of the accident and the publication of the recommendations is also highlighted. Railway accident investigation is a long process and might take years before reaching a conclusion. During this period of time, similar accidents caused by the same hazard might occur again. To mitigate such hazards effectively, proactively learning from other countries is critical to avoid such situations. The following sections offer case studies applying the techniques mentioned to real-world cases.

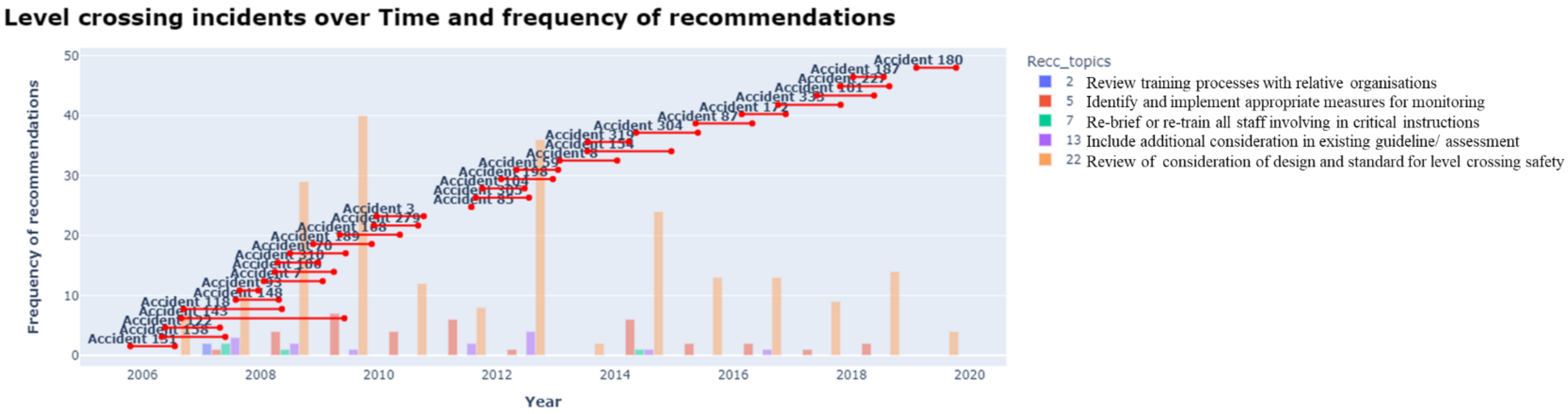


Figure 6-5: Level crossing incidents over time and frequency of recommendations (RAIB)

6.5 Case study I – level crossing accidents

Level crossing accidents have been widely discussed in the literature and have significantly influenced railway safety for a long time (Adeolu et al., 2016; Blaho et al., 2020; Bureika et al., 2018; Liang et al., 2018; Salmon et al., 2013). However, cross-country analysis is seldom found in the literature. This case study provides a comprehensive view of how hazards relevant to level crossing accidents impact the railway system across investigated countries and solutions made by each railway accident investigation body.

First, the *HazardMap* of each country for level crossing accidents is identified by developing the distribution of the relationship between hazards derived from the BERTopic model. Next, the threshold of covariance is set based on each distribution, which is set to 0.5 for the RAIB, NTSB and TSB datasets and 0.7 for the ATSB dataset. Once the distribution is generated, clusters relevant to the level crossing hazards are extracted manually. It is suggested to search relevant topics by starting with the top 50 topics and identifying the initial network. The network is further extended by looking at each document's topic distribution mentioning topics in the initial network. Note that the network of interest might be connected to other clusters. Therefore, the boundary must be manually identified to exclude irrelevant topics.

At this step, the threshold of the mentioning rate needs to be set to determine whether one document belongs to this network. A higher threshold results in a smaller number of selected documents with higher confidence of relativity and *vice versa*. To determine the best threshold of mentioning rate for each dataset, an initial rate can be set and documents with a mentioning rate close to the threshold should be manually reviewed. The threshold can be enlarged once most reviewed documents are irrelevant to the topic of interest and *vice versa*. Once relevant documents are retrieved, additional topics of interest can be further extracted to extend the network by reviewing dominant topics in documents.

An example of the extracted network for level crossing hazards in the NTSB dataset is demonstrated (Figure 6-6). The topic selected for identifying the initial network is topic 10: private crossings. Subsequently, the boundary is set after reviewing the relevance of topics on the edge as the initial network is connected to other clusters. Next, an initial threshold of the mentioning rate is set to 20%, and documents on the edge are reviewed. A final threshold is set to 10% and 36 documents are identified and labelled as level crossing (LC)-related incidents. Another relevant cluster containing two topics (topics 85 and 225) is also recognised after reviewing dominant topics in documents retrieved from the initial network.

After establishing the distribution of topics and their relationship relevant to level crossing accidents of each country, heterogenous terminology used in each country is standardised with the developed ontology (Section 5.4). For instance, mentions of "level crossing" and "grade crossing" are linked to and presented as the entity "level crossing (Q171448)". Topics with standardised names from investigated countries are clustered again based on the characteristics of hazards. Last, the *HazardMap* can be created by plotting hazards from each country with different colours for representations.

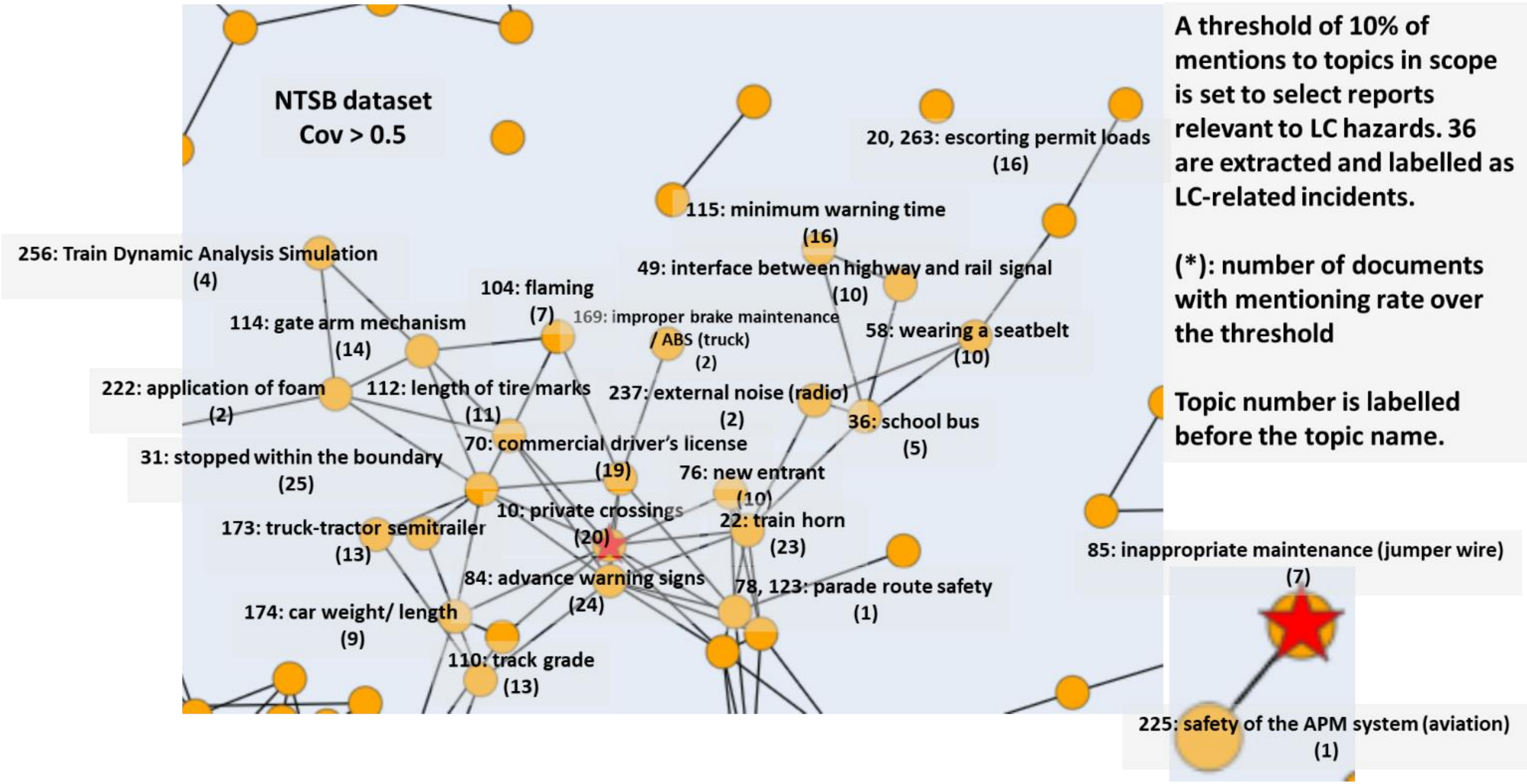


Figure 6-6: Distribution of topics and their relationship relevant to level crossing accidents published by the NTSB

There are 10 main hazards identified in *HazardMap* relevant to level crossing accidents investigated by national railway accident investigators: level crossing design (Figure 6-7), human factors, types of users, types of level crossings, external hazards, maintenance, and others (Figure 6-8), policy/management, employee training, and level crossing user education (Figure 6-9). Aspects of each hazard are coloured in accordance with identified countries. Overall, a significant number of aspects are observed in several hazards, including road signs, road users and policy/management. The RAIB and ATSB cover almost all aspects of hazards, and the NTSB places much emphasis on human factors. On the other hand, the TSB concentrates on types of users but overlooks the design of signs on the road and rail. The NTSB also investigated several potential behaviours of road users, such as stopping within the boundary and the regulation of users, whereas the ATSB focuses on the potential impact of the design of road signs and the condition of sighting distance. Thus, the difference in the approach that each country uses to address level crossing hazards between countries can be explored.

Finally, a cross-validation of the level crossing case study is conducted by using the Australian Level Crossing Assessment Model (ALCAM). ALCAM is an assessment system for identifying potential risks related to level crossing systems in Australia and prioritising the upgrade of dangerous level crossings by evaluating each level crossing with risky factors. Factors used in the ALCAM are extracted to conduct the comparison with aspects and hazards in the *HazardMap*.

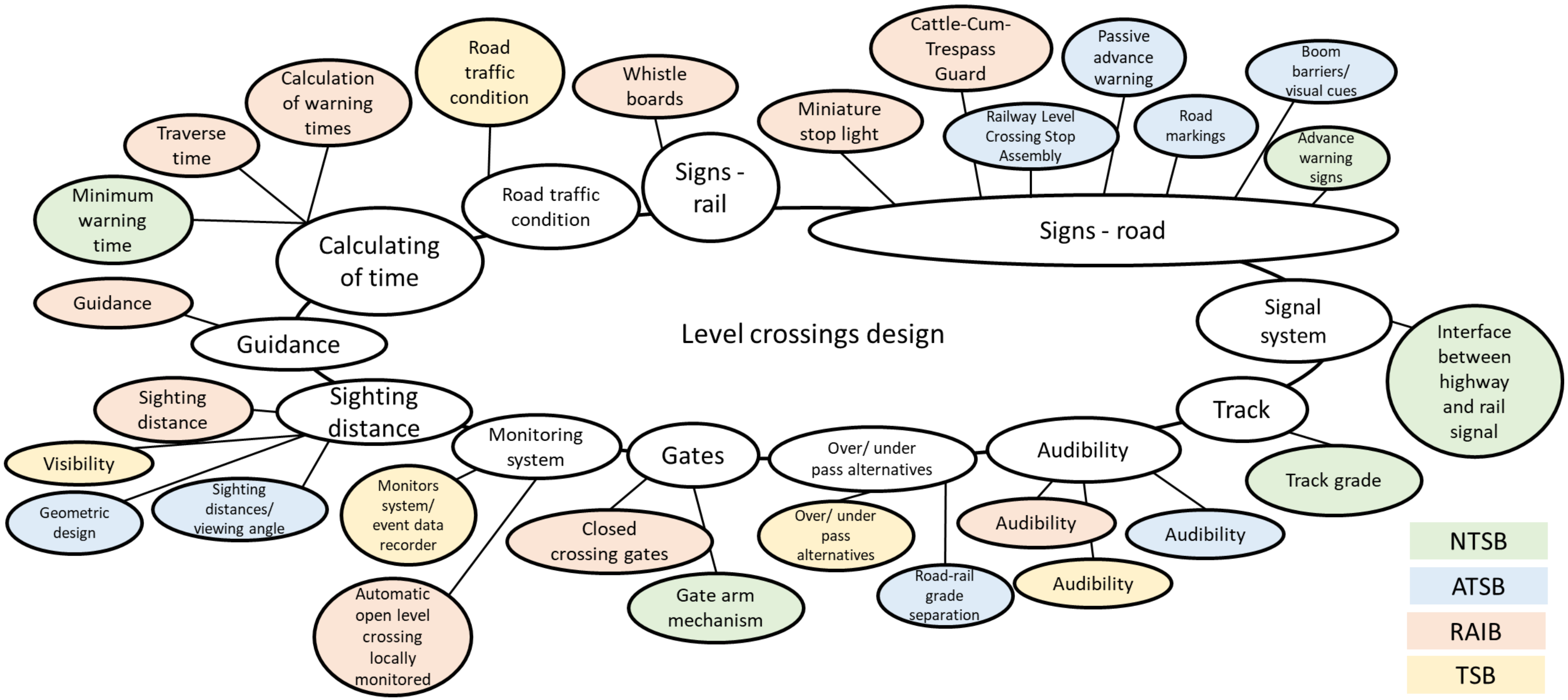


Figure 6-7: The applied HazardMap on level crossing accidents from four investigators – level crossing design

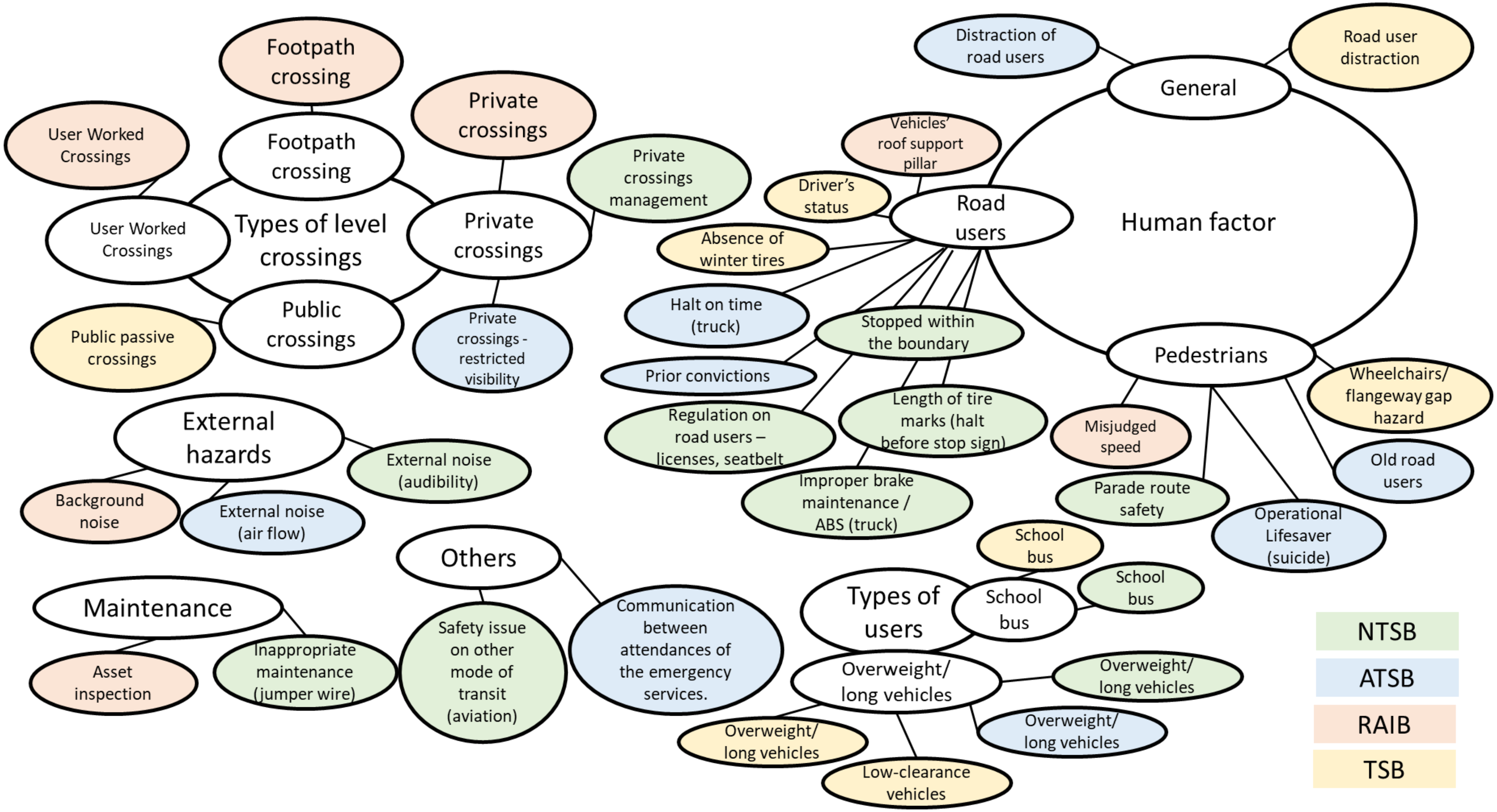


Figure 6-8: The applied HazardMap on level crossing accidents from four investigators – human factors, types of level crossings, external hazards, maintenance, types of users and others

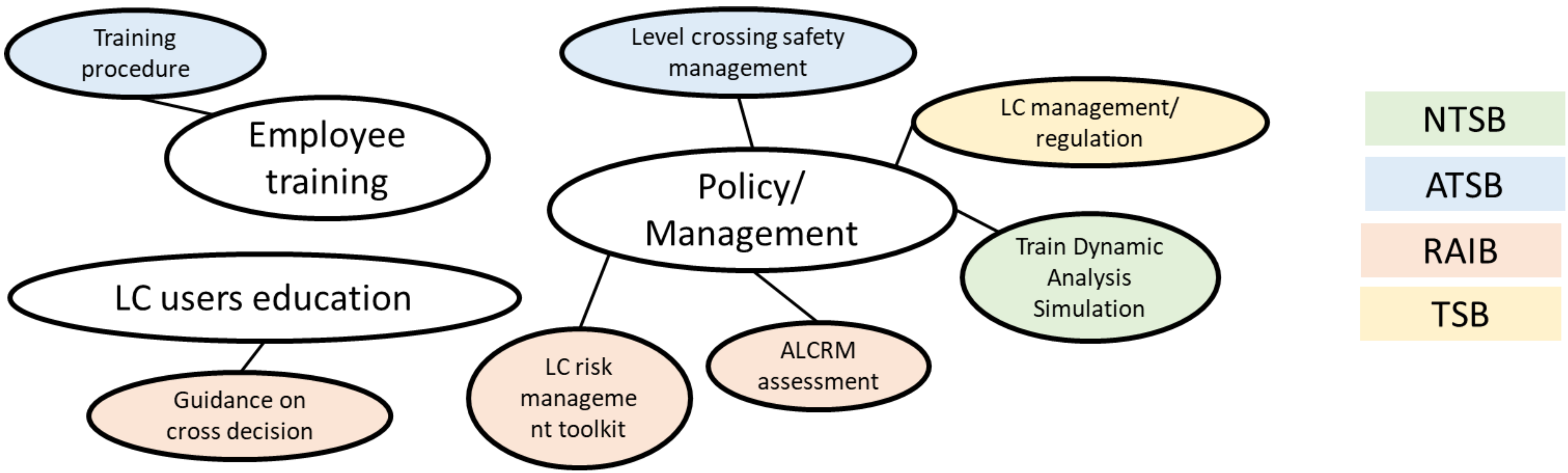


Figure 6-9: The applied HazardMap on level crossing accidents from four investigators – policy/management, employee training, and level crossing user education

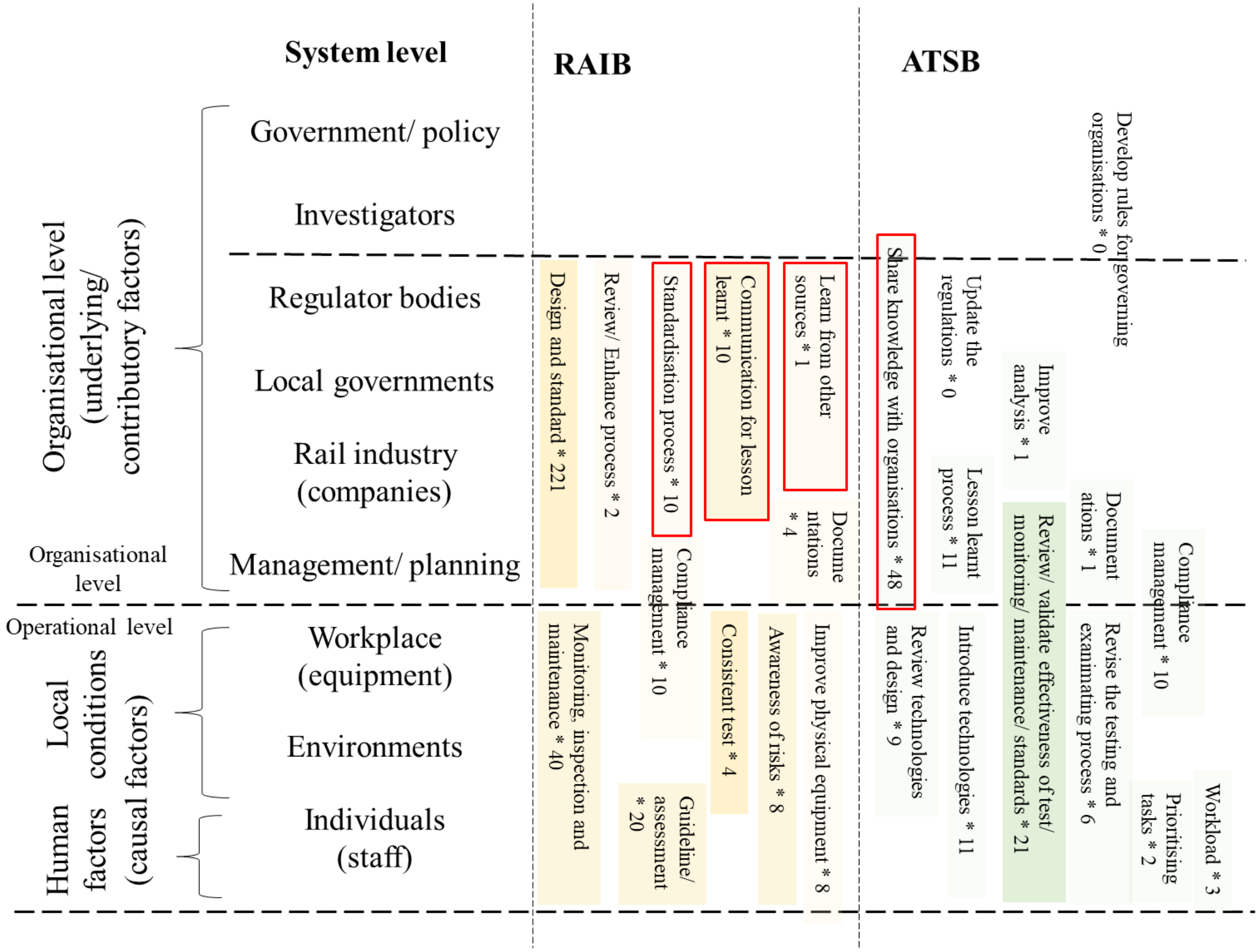
Overall, ALCAM focuses on distractions of road users, road and sign designs, types of road users and types of level crossing in detail. Results from our model cover almost all topics in the ALCAM but miss the following: proximity to sites or public facilities, the likelihood of vandalism to controls, seasonal/infrequent train patterns, train speed, train schedule and possible sun glare sighting. On the other hand, ALCAM places less emphasis on the conditions of road users, such as regulation on road users, improper vehicle brake maintenance and the absence of the use of winter tires. Suicide/trespass prevention and communication with emergency services are not included in ALCAM. Note that the lack of a characteristic in our model means the relation between this characteristic and level crossing is not significant from the analysis but may be substantial with other accidents. For instance, the connection between fatigue (road users and train drivers) and level crossing accidents is not found but the link with speeding is found.

On the other hand, *RecoMap* can be developed by extracting recommendations for level crossing accidents and adopting the process described (Section 6.2). However, the NTSB dataset is exclusive because the railway accident reports are inconsistent with the recommendation sources provided by the NTSB. Specifically, the codes of the recommendation dataset do not fully match codes assigned to railway accidents, hindering us from connecting the recommendations to corresponding reports. The TSB dataset is also excluded due to the insufficient amount of data.

Figure 6-10 shows the developed *RecoMap* on level crossing hazards based on recommendations proposed by the RAIB and ATSB. The RAIB has made a considerable number of recommendations relevant to the design and standardisation of level crossings, and an upward trend can be observed in recent years. Monitoring, inspection and maintenance also dominate the recommendations at the operational level. However, the number of recommendations related to learning and sharing knowledge is somewhat limited in the RAIB, so the level crossing hazards are recognised as local risks heavily relying on the context.

In contrast, the ATSB emphasises sharing knowledge with organisations in the railway industry, resulting in a series of independent railway level crossing safety groups and programs, such as the Regional Australia Level Crossing Safety Program. Therefore, a significant shift in the style of making recommendations to address level crossing risks is observed. Fewer recommendations have been made in recent years except for reviewing the effectiveness of the implemented programs and standards. Moreover, establishing independent railway level crossing safety management entities might be the ultimate solution made at the organisational level, providing high flexibility to collect data and conduct comprehensive analysis for addressing the risk as a whole.

Although it might be difficult to evaluate the different approaches that each country applies and determine the best and universal solutions to mitigate level crossing hazards, such risk has been drawing much attention and evidence has shown that overall level crossing risk has been mitigated significantly in recent years (Evans, 2011; Read et al., 2013; Salmon et al., 2016). However, understanding aspects of level crossing hazards and how they are addressed across jurisdictions helps maintain the level crossing risk within the affordable area and assists new railway transportation systems in developing a comprehensive framework to manage the identified hazards related to level crossings.



* Level of transparency represents the trend of recommendations made from the early years to the late years.
** The number followed by recommendations represents the occurrence of datapoints.
*** Recommendations in red outline represent the consideration of engagements with other organisations.

Figure 6-10: The developed RecoMap on level crossing hazards based on recommendations proposed by RAIB and ATSB

6.6 Case study II – the platform–train interface risk

National railway accident investigation bodies do not investigate all hazards or risks, resulting in a lack of analysable data. This circumstance might occur while processing data sources with a high level of heterogeneity. In this case, *HazardMap* and *RecoMap* might not be applicable. In addition to traditional qualitative analysis, the BERTopic analysis can be further extended to thoroughly understand how each identified aspect influences the railway system. A case study of the platform–train interface (PTI) risk is provided to demonstrate the details of such an extension. Note that additional functions are also applicable to the *HazardMap* depending on the research objective.

The PTI is a common hazard in the railway system. Although the consequences of such accidents might not be catastrophic or result in major disruptions, the PTI still causes a considerable number of casualties and injuries (Dai & Wang, 2010; Poirier et al., 2020). However, only a limited number of PTI accidents are investigated by the four countries at the national level. Both the ATSB and NTSB only published one PTI accident and the TSB has never published a report related to PTI. However, there is an individual program for addressing PTI risk funded by Australia and the US at the national level. An Australian government initiative published the report "Platform–train interface for rail passengers – a technology review" in 2012 (Devadoss et al., 2012), revealing potential hazards about PTI and proposing several solutions. On the other hand, another review was also conducted in 2015 by the Transit Cooperative Research Program (TCRP) Project A-40 to mitigate the PTI risks in the US. Despite the thorough review, both countries fail to continuously manage the PTI risk at the national level and monitor the implementation.

Conversely, the RAIB monitors PTI risks continuously as PTI-related accidents are still investigated. Despite a small number of reports, it is still worth examining them in a systematic way. Therefore, this section applies a modified analysis process to the RAIB dataset only.

Figure 6-11 shows the distribution of topics and their relationship relevant to PTI hazards in the RAIB dataset. First, the initial network for the PTI-related topics is extracted and reviewed by starting with topic 132: Platform edge gap. After reviewing reports on the edge a threshold of mentioning rate is set to 10%. Three irrelevant topics are removed from the initial network after investigating keywords of these topics (illustrated as the red line in Figure 6-11). Finally, a network consisting of 11 topics and 18 PTI-related reports is constructed. Additional relevant hazards are identified from reviewing PTI-related reports and shown in Figure 6-11. For the convenience of the following analysis, the topics in the network are referred to as the core PTI hazards, whereas additional relevant hazards are labelled as the supplementary PTI hazards.

Next, the measurement for understanding the extent to which each hazard influences the specific system should be designed. Discriminating the impact brought by aspects of the hazard on different railway systems helps decision makers to manage risks based on the characteristics of each system and mechanisms of historical accidents. Therefore, the covariate analysis elaborated in Chapter 4 is used to estimate the effect of each hazard on the selected variables as the proxy of the measurement.

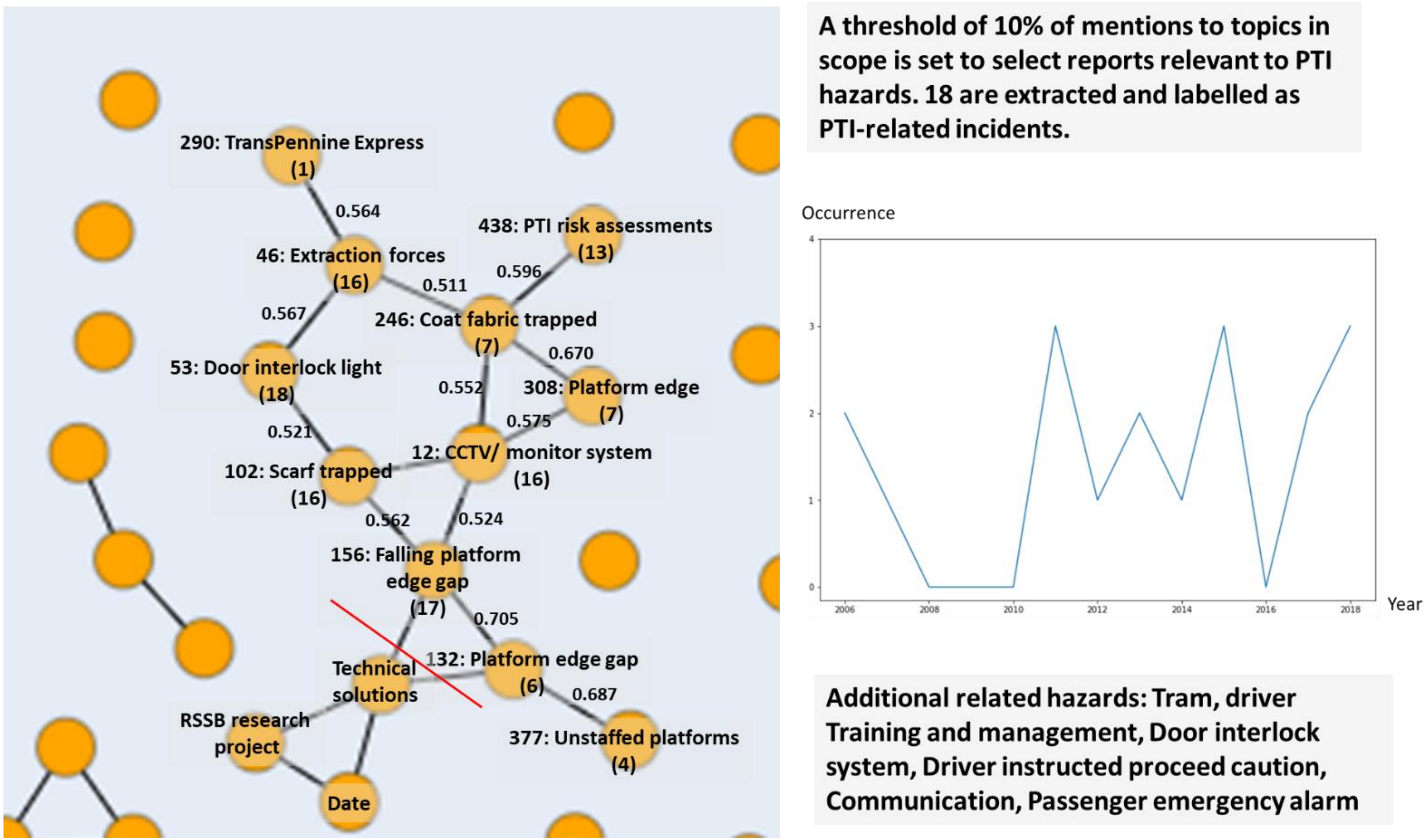


Figure 6-11: Distribution of topics and their relationship relevant to PTI hazards in the RAIB dataset

There is a significant number of topics connected to trams in the distribution of relationship between hazards identified in the RAIB dataset (Figure 6-3), implying that the tram is another mode of railway transportation that RAIB concentrates on. Thus, the covariate analysis is conducted by setting other modes of railway transportation as the fixed effect to measure how aspects of the PTI hazard cause tram accidents. Figure 6-12 shows the estimated effect[7] of core PTI hazards on trams against other modes from the RAIB dataset. Generally, the tram is more prone to all core PTI hazards. Scarf and coat fabric trapped dominate the causes of PTI accidents. Furthermore, the positive effect of the CCTV monitoring system and unstaffed platforms might reveal that a lack of platform security reduces the possibility of noticing trapped passengers and reflects the natural vulnerability of trams to PTI hazards. The (defective) door interlock light also indicates the potential combination of risks with human factors (not observing trapped passengers) to cause a PTI accident.

[7] The estimated effect refers to the proportion of each datapoint about a topic in an STM model estimated by a regression. This procedure involves measurement uncertainty from the STM model using the method of composition (Kwayu et al., 2021; Li et al., 2011; Roberts et al., 2019).

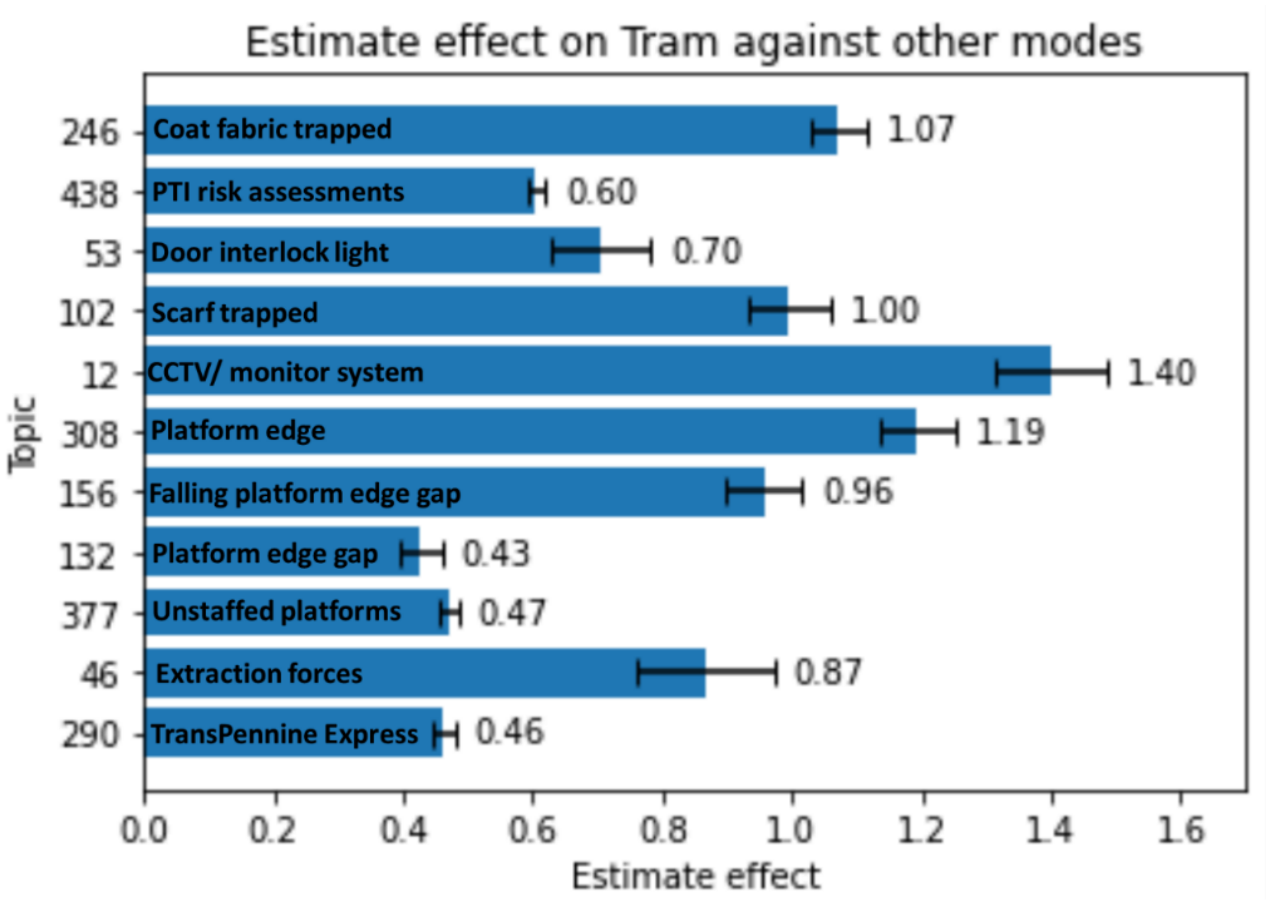


Figure 6-12: The estimate effect of core PTI hazards on trams against other modes from the RAIB dataset

On the other hand, another two positive effects on the tram relevant to the door interlock system are observed as well in the estimate effect of supplementary PTI hazards on trams against other modes from the RAIB dataset (Figure 6-13), namely “checking doors closed” and “door interlock system”. The topic “communication” has a slight negative effect on the tram compared with other modes of rail transport, implying that the operation of trams in the UK may heavily rely on the driver without other staff. Although there is no evidence that robust communication during operating the tram system can reduce the PTI-related hazard, the outcome still indicates that communication plays a vital role in such hazard regardless of the mode of the railway system. By adopting the estimate effect analysis and synthesising findings, decision-makers are able to address the hazard of interest on the basis of the characteristic of each railway transportation system and customise corresponding recommendations.

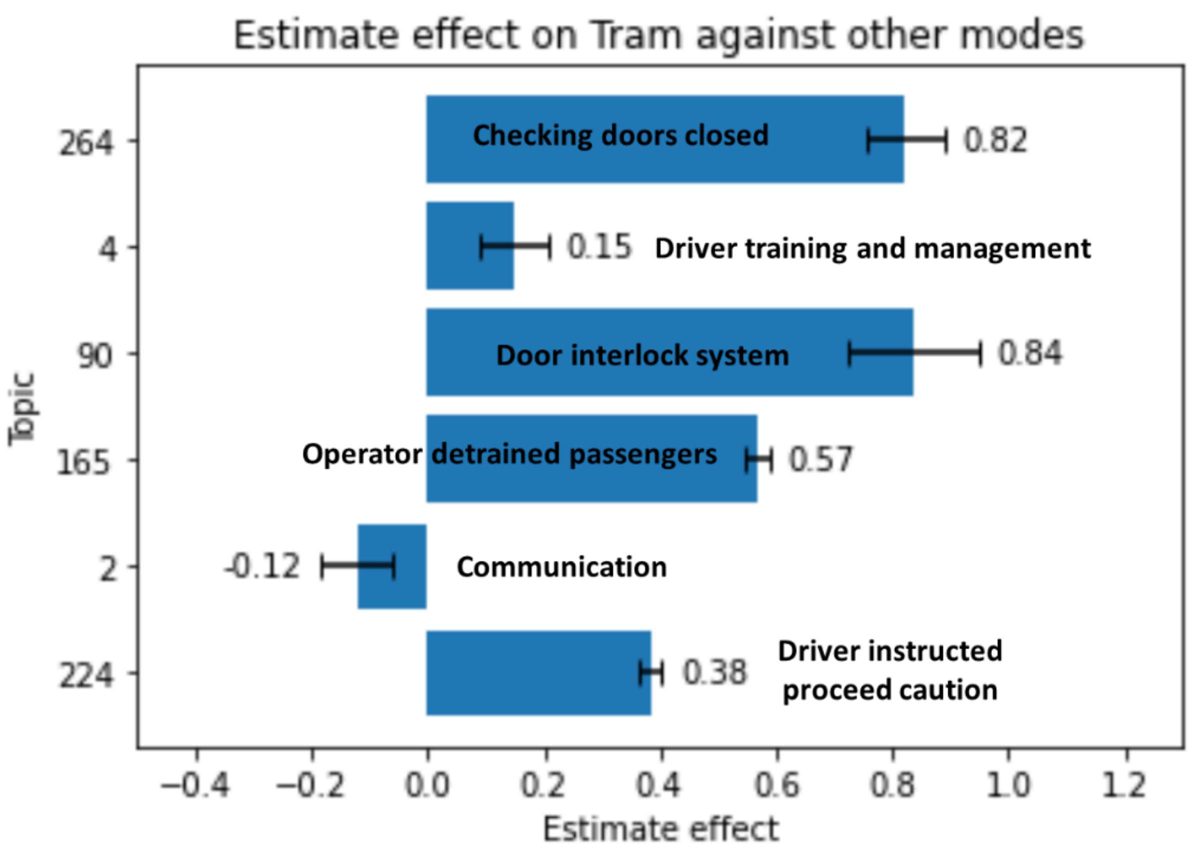


Figure 6-13: The estimate effect of supplementary PTI hazards on trams against other modes from the RAIB dataset

Last, the PTI-related incidents over time and the frequency of recommendations are shown in Figure 6-14. Overall, the number of accidents related to the PTI-related hazard has been increasing since 2011. The time for an investigation is 1 to 2 years. Most recommendations are made to improve the physical equipment or infrastructure. In addition, a growing trend of reviewing training processes can be observed from 2016. This might indicate that the RAIB first suggests the improvement of the infrastructure and equipment once the hazard is identified, followed by organising a thorough training system and procedures to mitigate the hazard of the interface between employees and introduced infrastructure. Such mixed approaches benefit the system by addressing different aspects of the PTI hazard from a variety of perspectives.

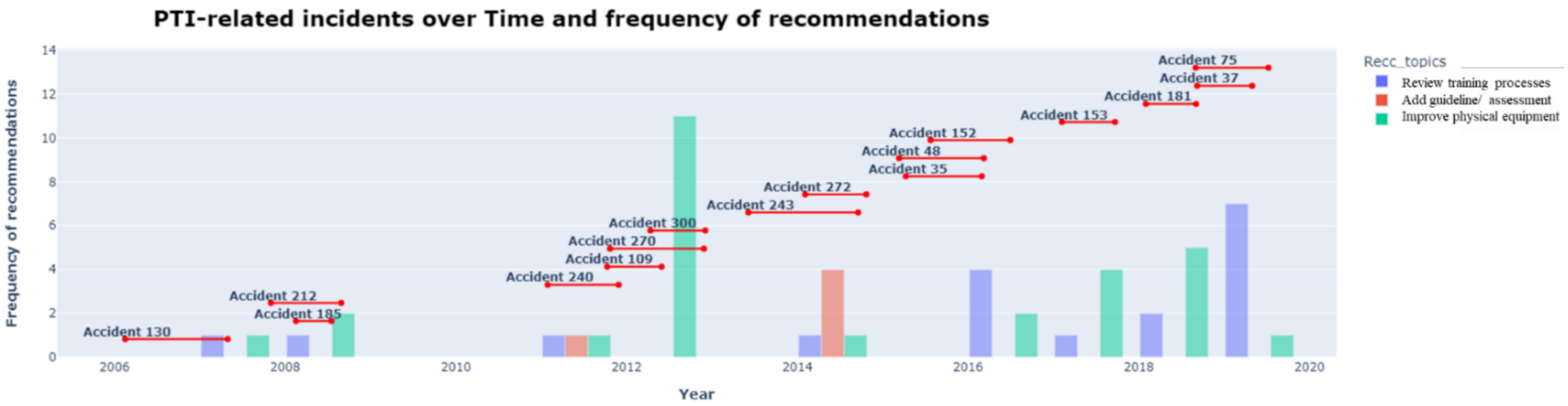


Figure 6-14: PTI-related incidents over time and frequency of recommendations (RAIB)

6.7 Synthesis of findings

The development of *HazardMap* and *RecoMap* introduced in this chapter has potentially provided solutions by further extending the raw outcome and incorporating it with theoretical frameworks in the literature. Practitioners were able to have a comprehensive view of how railway accidents are triggered in other jurisdictions and how hazards identified are addressed with recommendations proposed in different countries. Further connection between existing railway hazard taxonomy frameworks developed by each jurisdiction and *HazardMap* can be built to discover hazards potentially overlooked. Additionally, several learning behaviours in the railway industry were observed in the workshop sessions and survey, which are further discussed in the next chapter.

Cross-sectional analysis was conducted on railway accidents and investigators. It indicated that ignoring topics with a lower probability of occurrences from the topic model might result in missing critical mechanisms in railway accidents. Therefore, a systematic view of hazards is required to comprehensively understand how hazards trigger an accident. On the other hand, the cross-sectional analysis of investigators suggested that each body has various styles of making recommendations during different periods. However, a comprehensive overview of recommendations made in the past cannot be illustrated for comparison between countries by only analysing the outcome retrieved from the STM.

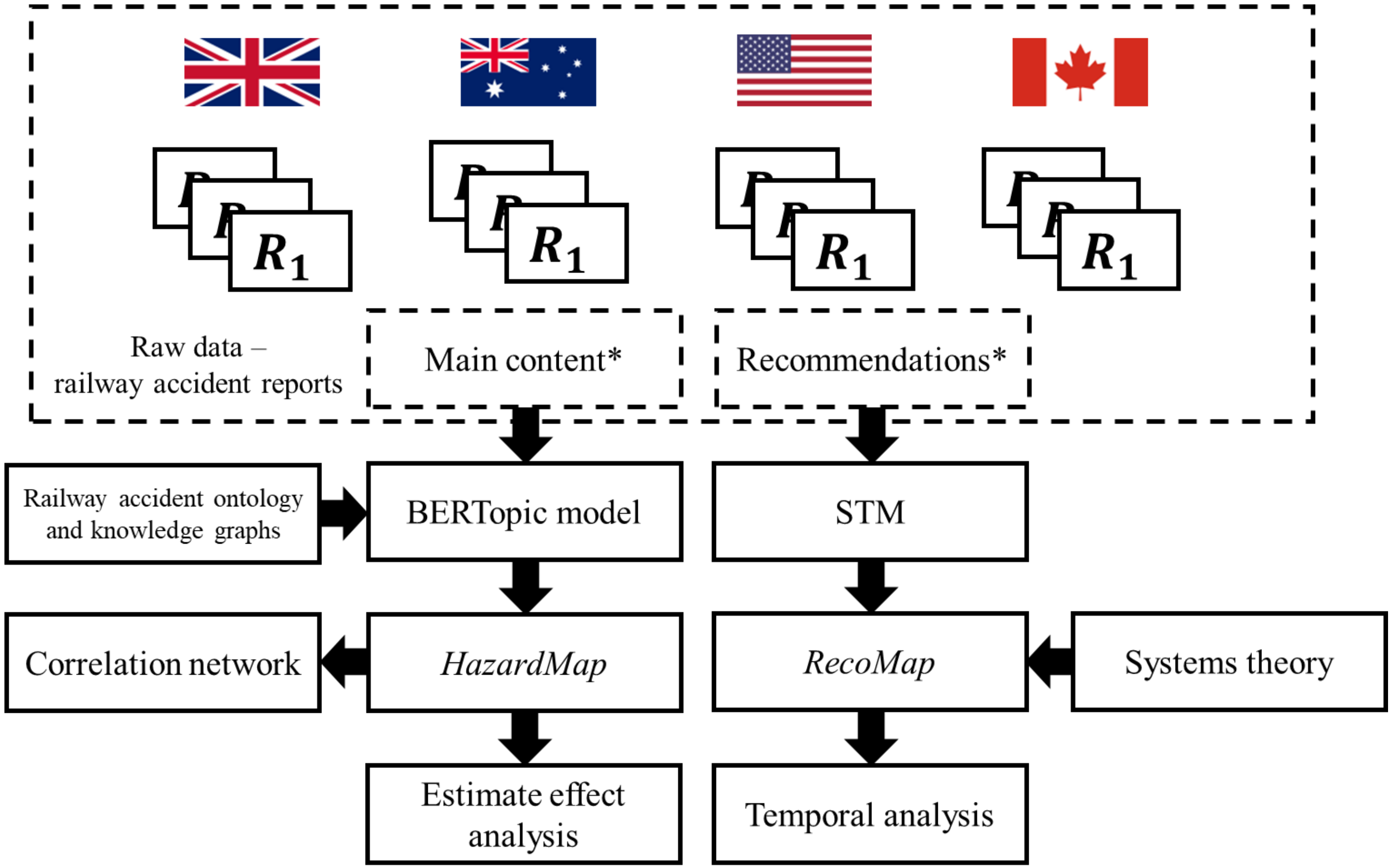
To overcome such limitations, *RecoMap* is proposed based on the outcome of the STM to understand the perspective by which each recommendation addresses the hazard. The trend of recommendation styles that each investigator makes can be revealed from the view of the socio-technical system. Several types of recommendations are introduced as well to discriminate the role each recommendation plays from the perspective of the railway industry. The result suggests that authorities at different levels should concentrate on solutions based on the jurisdiction. Recommendations with interfering instructions might only be appropriate for local authorities given that the cooperation with the local railway industry might be tighter than other organisations at the national level. On the other hand, individual railway accident investigators should emphasise advancing a railway safety culture by promoting engagement between organisations in the railway industry from different areas. Specifically, learning and sharing knowledge should be led by the national railway authority to ensure that hazards can be mitigated in advance by learning across jurisdictions and across time. Additionally, supportive recommendations should be made by the national railway authority given that each railway industry might have inherent characteristics of the operating system and require more flexibility to adopt general recommendations.

On the other hand, *HazardMap* was also proposed to depict the nature of hazards in the railway system and their mechanisms on different countries. *HazardMap* is developed based on the output of BERTopic, enabling the consideration of topics with a low probability of occurrences and the visualisation of the relations between hazards, accidents and recommendations across countries. *HazardMap* can also describe how each hazard triggers a railway accident by revealing the unaddressed aspects. Therefore, railway accident investigators in different countries can understand the potential pattern of how one hazard impacts the railway system by reviewing the mechanism found in other countries and developing corresponding solutions before it triggers another railway accident.

Last, two case studies were delivered to show how the proposed model can be applied under different circumstances. The temporal analysis and sensitivity analysis were demonstrated with actual cases. However, several limitations were identified while collecting the data in the case studies. For example, the TSB in Canada only proposes recommendations when the severity or the consequences of the railway accident significantly impact society, making the recommendations dataset relatively small. Another example is that the recommendation dataset the NTSB provides does not fully match the original railway accident reports, resulting in poor connections between the two data resources. These limitations might result in challenges in applying the proposed model comprehensively across different jurisdictions as the selective nature of railway accidents investigated by each country could cause imbalanced and potentially biased representations of railway safety issues. Such an inconsistency might hinder models from providing a universally applicable analysis. Fortunately, these limitations only influence the analysis of specific hazards given that the individual recommendation dataset of each country is comprehensive. But the issue of generalising and standardising the analysis across countries still needs to be addressed in future work.

Figure 6-15 shows the overview of the data flow and analysis procedures for developing *HazardMap* and *RecoMap*. The main content of railway accident reports was used to extract potential hazards for constructing the *HazardMap*. The ontology and knowledge graphs were applied to standardise the terminology used in different countries. The developed *HazardMap* is capable of offering an overview of the distribution of hazards identified from historical railway accidents in each country. Furthermore, the advanced application of the *HazardMap* can help users to understand the relationship between hazards and the extent to which each hazard impacts the specific railway system via the correlation network and estimate effect analysis respectively. On the other hand, recommendations made in railway accident reports are extracted and analysed separately to understand how each investigation body addresses risks identified. *RecoMap* was developed based on the results of the STM and systems theory, enabling users to understand the distribution of recommendations made on the socio-technical hierarchy. *RecoMap* was further extended for temporal analysis, providing additional insights of solutions during different periods of time. To sum up, this work contributes to the railway safety context by offering the opportunity to examine a large volume of railway accidents from multiple perspectives and allowing end users to have a comprehensive view of hazards across countries and across time.

Although a railway safety culture has been promoted for several years and there has been a significant reduction in fatalities and injuries of recent railway accidents in many countries, keeping improving and maintaining the objective of safety in a dynamic environment is still a significant challenge for the railway industry around the world. Learning across jurisdictions and across time is likely to continue to be essential given that due to changing environments such as severe climate change a railway system is likely to encounter hazards unknown to it but that have been fully addressed in other countries. The model and case studies proposed in this study can help the railway industry overcome the difficulties of learning across jurisdictions and across time and help enhance the understanding of railway safety. High similarities of hazards identified in the case studies (Section 6.5 and 6.6) and limited learning behaviours across jurisdictions (Section 5.5.2 and 5.6.2) suggested that railway accidents are likely to repeat themselves across jurisdictions unless the lesson learned behaviour is enabled. Several barriers to learning across jurisdictions were also identified (Section 5.6.2), implying that the poor motivation of learning behaviours at the organisational level. Such systematic issues might cause a potentially emerging hazard deteriorating railway safety culture. Therefore, the next chapter elaborates on the railway safety deterioration as another critical hazard among railway industries across the world.



* Railway accident reports are divided into main content (including all descriptions such as summary, investigation process and conclusion) and recommendations for different purposes of analysis. More details can be found in Chapter 3.

Figure 6-15: Overview of the data flow and analysis procedures for developing HazardMap and RecoMap

7. Deterioration in railway safety: a potentially emerging hazard

This chapter illustrates the potentially emerging hazard of deterioration in railway safety from current stable railway systems worldwide and the possible pattern of how systems become hazardous. Potential solutions for overcoming the barriers with proposed models and further work suggested are discussed in detail. The structure of the chapter is as follows: first, the learning behaviours in the railway industry are discussed (Section 7.1), followed by the analysis of railway safety knowledge retrieving, processing and disseminating (Section 7.2). Another underlying hazard, the potential deterioration of railway safety culture, is revealed and discussed (Section 7.3) along with the opportunity to overcome barriers (Section 7.4). Finally, the synthesis of findings is presented (Section 7.5).

7.1 Learning behaviours in the railway industry

Observing the learning behaviours in the railway industry ensures that lessons can be thoroughly learned, and hazards can be identified, prevented and managed before they trigger similar accidents. Although the result of the case studies (Sections 6.5 and 6.6) suggests that similar accidents might still occur over time, the *HazardMap* discussed in Section 6.3 has shown that such circumstances happen because different aspects of the hazard trigger the accident in combination with other hazards and cause similar patterns and consequences. Additionally, several hazard aspects appear to have triggered an accident and have been discussed in various jurisdictions. For example, audibility has been found to play a critical role in the level crossing hazard by the RAIB, ATSB and TSB (Figure 6-7 to Figure 6-9), implying homogenous aspects of a hazard can be found in one type of accident regardless of the investigated locations. On the other hand, some aspects of a hazard have only triggered a limited number of accidents indicating that such seen aspects might still be underlying unseen factors in other jurisdictions and might trigger another railway accident.

Therefore, the priority strategy for mitigating hazards is to learn from other jurisdictions and reveal these underlying unseen factors that have been reviewed and discussed in other locations.

According to the results of Chapter 5, several observations potentially indicate the learning behaviours in the railway industry. First, the frequency of co-reference between RAIB, NTSB, TSB and ATSB (Figure 5-30) shows a high level of self-reference, implying learning behaviour across time given that reviewing similar railway accidents is a component during the investigation. However, the reference to other external sources by an individual jurisdiction is relatively lower, which might indicate the restricted learning behaviour across jurisdictions. Second, the *RecoMap* has revealed a small number of recommendations supporting learning behaviours across organisations such as learning from other sources and sharing and disseminating knowledge. This implies the potential motivation of railway accident investigators to promote learning behaviours at the organisational level to mitigate systemic railway accidents. However, such promotion seems to be mostly restricted to local jurisdictions, indicating less consideration of learning from external resources from other countries.

Figure 7-1 shows the information flow between stakeholders in the railway industry. Note that this illustration assumes that the information related to railway safety primarily comes from railway accidents and that railway accident investigation bodies play the major role in promoting railway safety and learning behaviours in one jurisdiction. According to Figure 7-1, a strong learning behaviour across organisations within one country can be found, which has also been supported by the literature (Akel et al., 2022; Paul et al., 2018). Both railway accident investigation bodies and organisations in one jurisdiction proactively learn from (previous) railway accidents by conducting investigations and analysis to mitigate hazards identified and enhance railway safety.

Meanwhile, previous railway accidents are retrieved to support the analysis of the current railway accident and understand what implementation should have been in place to mitigate hazards revealed. This information flow shows the action of learning across time and organisational learning (more details are given in the following sections). However, the information exchange across jurisdictions in the railway industry, which has been observed from the frequency of co-reference between RAIB, NTSB, TSB and ATSB (Figure 5-30) and the analysis of the scoping workshops and survey, is limited. Such a gap might implicitly result in another underlying risk of failing to manage aspects of hazards identified in other areas before they trigger an accident. To discover the potential barriers and incentives to learning across jurisdictions in the railway industry, we first investigate how railway safety knowledge is retrieved, processed and disseminated in complex adaptative systems.

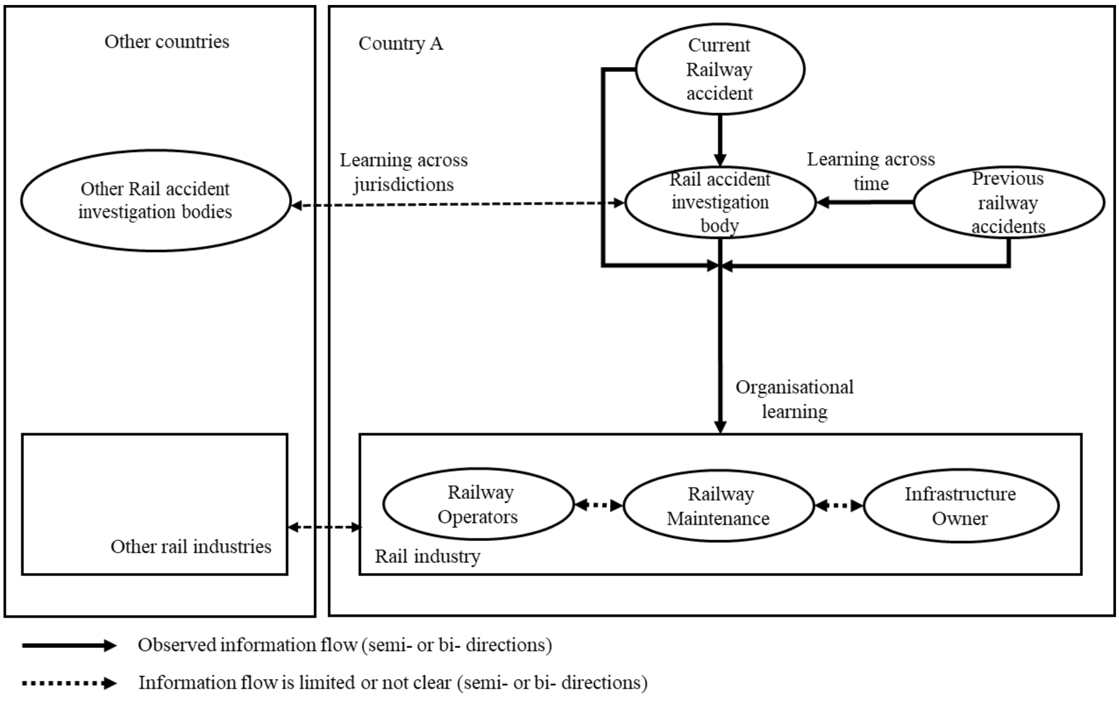


Figure 7-1: The information flow between stakeholders in the railway industry

7.2 Railway safety knowledge retrieving, processing and disseminating

For the information exchange between organisations and learning behaviour, most studies in the literature focus on staff mentoring (Holmes & Robertson, 2021; McHugh & Klockner, 2020; Naweed & Ambrosetti, 2015) and the promotion of safety culture in an individual organisation (Placencia, 2016; Wilson & Norris, 2005). However, institutional interplay in organisational learning is rarely discussed in the railway context.

From the perspective of knowledge flow, the resource-based view and inter-firm network theory suggest that firms need to build and manage networks to retrieve knowledge and produce information of value through using internal capabilities (Huggins & Johnston, 2010). Considering that railway safety is a product managed by organisations in the railway industry and the learning behaviour can be considered as the information and knowledge transmission at organisational level, then the amount of safety-related knowledge flow, including receiving, possessing and transmitting, would determine how the railway safety culture is formed (Huang et al., 2019; Johnsen et al., 2006). Therefore, this argument may have supported the relationship between information flow and the formation of the safety culture. More details are illustrated in Section 6.3.

Several works in other contexts such as nuclear power and the healthcare industry show a strong interest in organisational interplay on learning behaviour. Duffield and Whitty (2015) proposed the “systemic lessons learned knowledge model” (Figure 7-2) and revealed the phenomenon that organisations seldom learn from prior experience while the models and guides of implementation are transparent and available. Several facilitators are proposed by Duffield and Whitty (2015) for practical implementation. For example, social facilitators can acknowledge individual, group and team activities and rewarding works. Cultural facilitators include valuing and encouraging people to contribute, providing support to those who want to increase their knowledge and regularly updating on the organisation focus. Other works relevant to this context have been discussed and adapted to several practical application areas (Cedergren & Petersen, 2011; Filho et al., 2021; McHugh & Klockner, 2020).

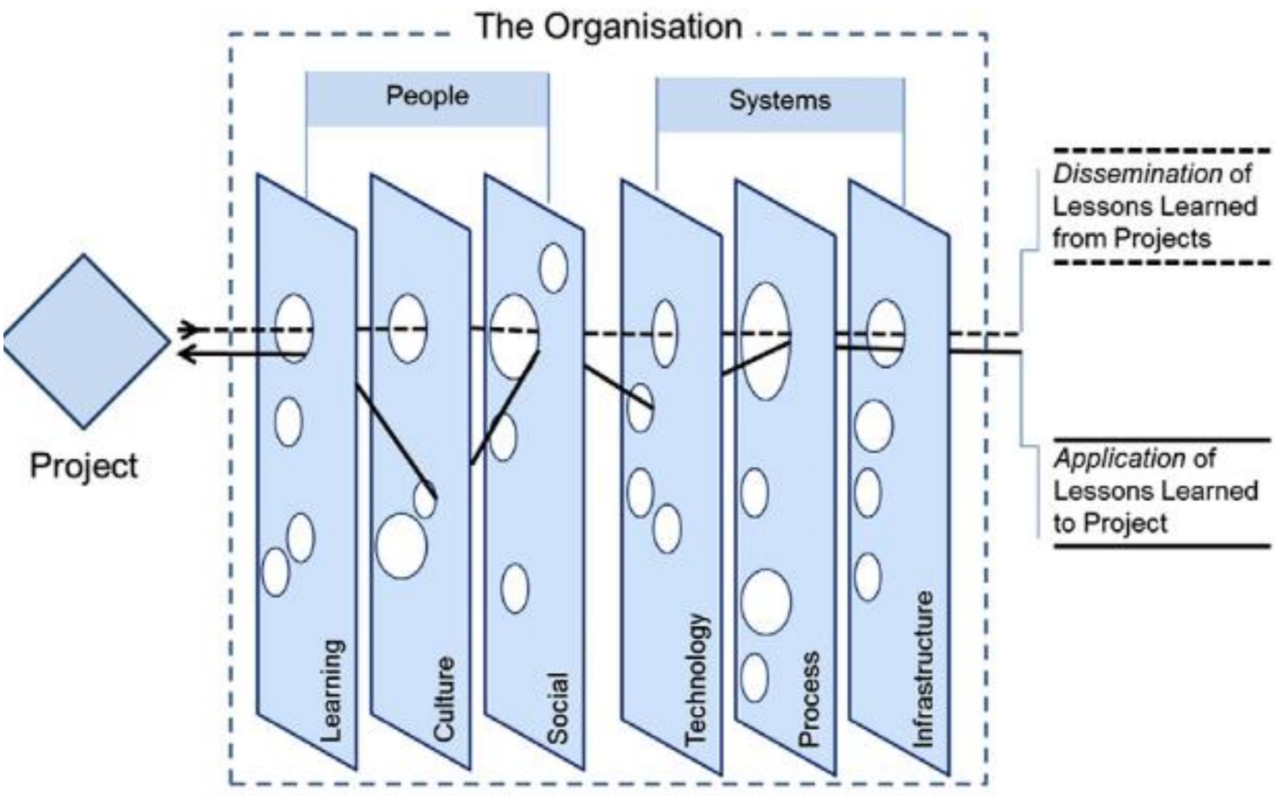


Figure 7-2: The systemic lessons learned knowledge model (Duffield & Whitty, 2015)

In terms of potential barriers and incentives of organisational learning from the perspective of the industry, Størseth and Tinmannsvik (2012) found that organisations would be less willing to learn if 1) they are forced to do so, 2) the media influences the focus, 3) they are asked to provide instant responses by carrying out various hurried but unorganised actions, 4) the trend of safety becomes void and irrelevant to operation, and 5) the absence of procedure has been proven, making that a way to improve. On the other hand, potential incentives to learn from peers are 1) there is no blame culture, 2) no urge for procedures to be released, and 3) there are well-organised documentation systems and adequate learning skills. In addition to the detailed barriers and incentives discussed by Duffield and Whitty (2015), prior works show that complex adaptative organisations fail to incorporate application and dissemination (Duffield & Whitty, 2015). In other words, proactive dissemination and application of lessons learned are absent although lessons (accidents) are successfully identified, implying that the underlying motivation driving a complex adaptative system to learn might be more complicated at the systematic level.

The term complex adaptative system describes a system that forms its behaviours based on prior experience and knowledge embedded in previous environments or conditions. Such a process enables this system to react to one incident under a particular environment based on previous cases with similar conditions (Duffield & Whitty, 2015). Several important industries adopt the complex adaptative system, including the nuclear power industry, healthcare organisations, medical systems and large-scale public transport systems such as aviation and railways. The learning process in the complex adaptative system might be driven by a variety of motivations and be based on the characteristic and purpose of the system. For instance, the lessons learned for practice in the nuclear power industry are maintained by the Institute of Nuclear Power Operations, aiming to standardise the procedure of precursors identification and knowledge dissemination. Aviation has a similar power-centralised international regulation body.

In addition to complex adaptative systems, the healthcare industry also adopts the high reliability organisation concept via thorough communication, peer review, peer coaching, team behaviour, reporting and systematic analysis. When an incident occurs, the hospital in charge works with academia and informs the World Health Organization once underlying factors discovered in the incident are found to potentially cause damage in other countries. These processes are implemented in the accident information collection, analysis and dissemination to ensure critical hazards can be identified, proactively mitigated, and disseminated. Such a peer-review-based process keeps the motivation of advancing safety culture through learning across jurisdictions.

Furthermore, broader cross-sector learning has been gradually proposed in the chemical industry since the occurrence of several catastrophic accidents. A review of 15 years of lessons learned reveals that ineffective information dissemination due to the lack of institutional support significantly reduces the reliability level (Hedlund et al., 2014). Therefore, learning plays a critical role in committing to a safe organisational environment in the railway industry (Jeffcott et al., 2006).

In work on aviation safety, the first significant improvement was attributed to novel technology. Subsequently, human factors are frequently discussed, concentrating on the solution of the human–machine interface. In recent years, the focus has gradually shifted to the impact of organisational factors such as policies, norms and other latent factors that might influence safety performance (Patankar & Sabin, 2010). All knowledge abovementioned is integrated by the International Civil Aviation Organisation (ICAO) as the basis of international aviation operation regulation. Therefore, organisations cooperate and share information under the scheme to maintain the best aviation safety practice.

According to materials retrieved from the roundtable discussion in the scoping workshops and survey conducted as part of this research (discussed in Section 5.6.3), standardised jurisdiction-based analysis dominates the process of advancing railway safety. Learnings from historical accidents have been converted into knowledge and experience now embedded in day-to-day operation. However, cross-jurisdiction-based analysis has not been fully incorporated into the process due to a series of barriers, such as technological limitations and the lack of motivation. Similar circumstances have been identified in the *RecoMap* given that the concept related to cross-jurisdiction-based analysis is absent. The potential weak indication has also been revealed because several workshop and survey respondents mentioned that learning across jurisdictions only happens in connection with specific hazards and specific investigators. Even though there is an existing local authority responsible for railway safety in one jurisdiction, information and knowledge can barely flow between countries due to a lack of centralised authorities. This means that the proactive lesson learned process can only be found on a local rather than global scale.

Figure 7-3 illustrates the information flow integrated with *RecoMap*, workshops and survey results. Apart from the use of theories and limited external sources mentioned above, additional observations indicate several practitioners' concerns about behaviour. It should be noted that no participant in the workshops and survey represented railway accident investigators. Therefore, the proposed argument only concentrated on practitioners' attitudes and the potential link to the formation of a railway safety culture.

First, the concern about the legislative framework has been raised in many fields such as engagement with other organisations, supporting railway accident investigation bodies and disseminating data and knowledge. This phenomenon may be derived from the history of the railway industry. Several works have revealed that proper legislation can increase the quality of safety and address human factor issues (Burdzik et al., 2017; Priestley & Lee, 2008; Railtrack & House, 1997). Hence, this concept has become a paradigm followed by researchers and practitioners in the railway industry and has resulted in many management approaches, such as the "just" culture (Clarke & Clarke, 2000; Hutchings & Thatcher, 2019; Naweed et al., 2022). However, this may result in a conservative and centralised culture (Lingegård & Lindahl, 2015), which has been revealed in the scoping workshops and surveys. The railway industry also believes that such a legislative framework and rigorous standardisation can lead to a safe environment once the risk has been mitigated to "as low as reasonably possible" (ALARP) and might have overlooked the potential impact of dynamic external and internal factors on the complicated system.

Second, a robust knowledge extraction process can be observed in Figure 7-3. The railway industry in each jurisdiction puts much emphasis on the systematic analysis of each railway accident such as applying theoretical frameworks, discovering underlying factors, and collecting internal and external evidence. However, several considerations might have hindered the railway industry from advancing learning behaviours through accident analysis. For instance, the cost and expected performance on improving railway safety increase uncertainty during decision-making. Nevertheless, practitioners also suffer from the lack of adequate technology to assist the analysis process, limiting the effectiveness of the learning process. The legislation also plays a critical role in accident analysis given that materials related to railway accidents might be sensitive to society. Thus, processing and disseminating information has been primarily restricted to legislative frameworks.

However, the European Union Agency for Railways (ERA)[8] does play a role as a centralised authority responsible for setting mandatory requirements and standards for European railways and manufacturers, concentrating on the technical interoperability between systems. The major responsibility of ERA is to offer technical support to increase the efficiency in manufacturing vehicles and monitor national safety rules, safety performance, and the progress towards interoperability within the EU. Despite the critical role that ERA plays across countries in the EU, the regulation authority is still restricted to the national level by individual members. Thus, ERA is unable to determine the mandatory operational standards and only partly influences the development of national regulations while providing guidelines and recommendations to improve safety. Given that ERA has sufficient knowledge and coordination expertise in railway safety across countries in the EU, it could potentially become the ideal body in promoting cross-jurisdiction learning through novel technologies.

Last, it is worth noting that the railway industry in the UK has established an individual organisation called the Rail Safety and Standards Board (RSSB), representing the aggregation of members of the railway industry in the UK. It is responsible for improving

[8] https://www.era.europa.eu

railway safety and leading research and projects to address recommendations made by RAIB. In other words, the Rail Safety and Standards Board is the proxy of practitioners in the railway industry such as Network Rail (the infrastructure operator), rolling stock companies and train operating companies. The Rail Safety and Standards Board is also responsible for information retrieving, processing and sharing to ensure members can implement safety-related actions consistently without devoting additional effort to conducting analysis. Thus, a representative entity such as the Rail Safety and Standards Board can be considered as another type of local centralised authority representing the railway industry and providing a strong incentive to share information.

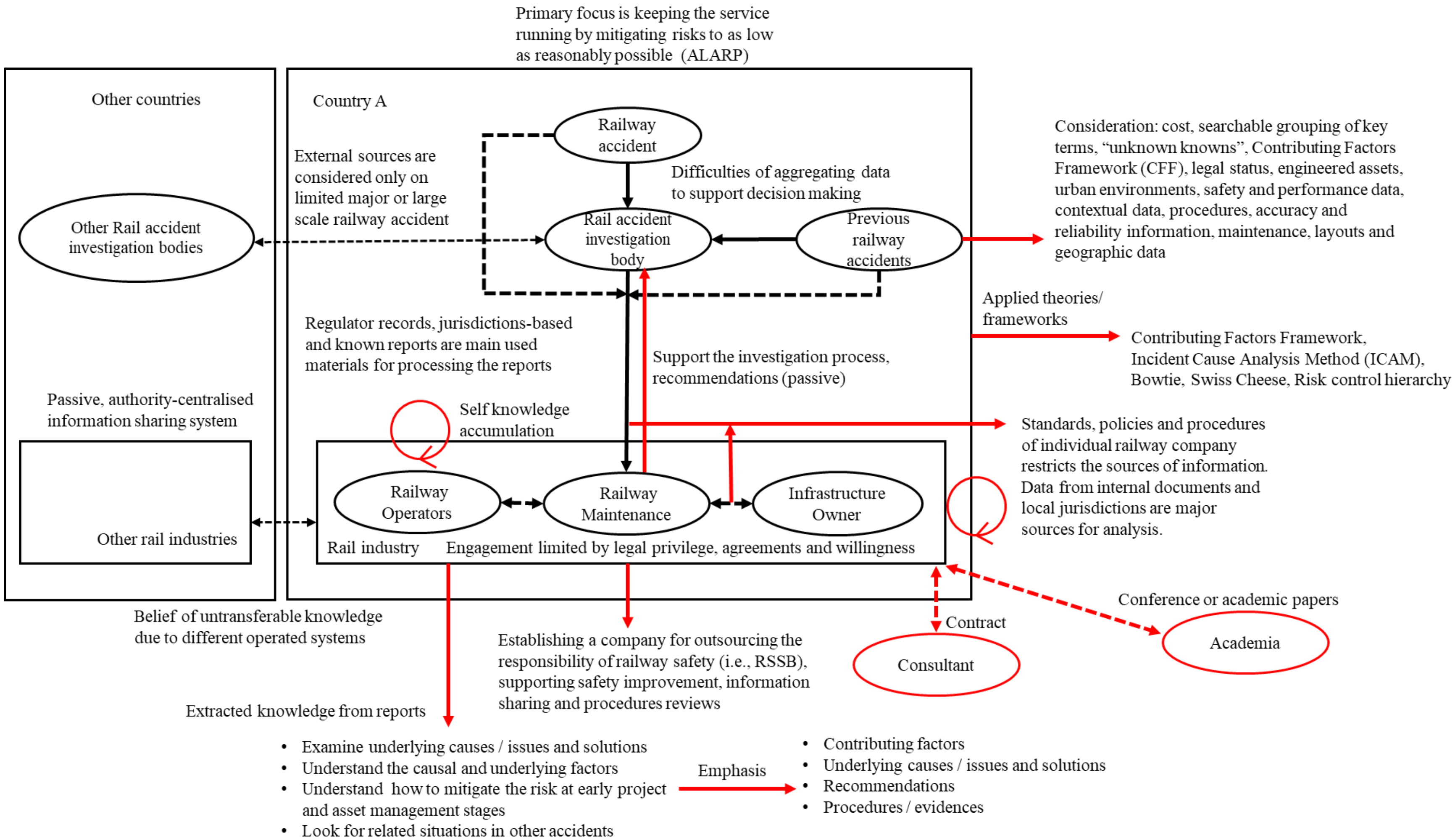


Figure 7-3: Existing information flow integrated with the results of RecoMap, workshops and survey

7.3 Another underlying hazard: the potential deterioration of railway safety culture

Previous sections of this chapter indicate the potential barriers. Incentives to learning across jurisdictions and across time are also discussed. Safety culture is the underlying motivation driving an industry to commit to a safe environment or not. The theoretical definition of the safety culture is "a pattern of shared basic assumptions that the group learned as it solved its problems of external adaptation and internal integration, that has worked well enough to be considered valid and, therefore, to be taught to new members as the correct way to perceive, think and feel in relation to those problems" (Schein, 1992, p. 12). According to the definition, three core concepts can be identified: basis assumptions, espoused values and artefacts. The basic assumptions are the implicit assumptions guiding the behaviour of group members and mitigating the variation in a cultural unit in terms of how to observe, think and feel about things. The basic assumptions are conceptual ideas and can only be observed through artefacts and espoused values. On the other hand, the espoused values are explicit attitudes that manifest the organisational climate regarding hardware, software, people and risks. Most espoused values can be observed directly from policies, training, manuals or procedures. Last, the artefacts are visible objects or activities subject to espoused values, such as statements, meetings, inspection reports and equipment. Despite the visibility, the artefacts cannot reveal the underlying culture of an organisation. Safety culture development has been found to rely on employee learning from reviewing outcomes, incentive-enabling factors and consistency over time (Bisbey et al., 2021). It is also suggested to maintain conducive conditions for safe behaviour norms, values and core beliefs to develop collective safety culture (Nordin et al., 2020). Maintaining the culture of safety is the task necessary to avoid the potential decline in awareness over time while achieving a safe work environment and has been integrated with process standardisation, protocol development and team training.

Most studies in the literature concentrate on transforming safety culture at the organisational level toward a safer working environment; for example, changing from a blame culture to a just culture (Whittingham, 2012; Wittmer & Mark, 2021). The focus is primarily on proactively developing a better safety culture and what factors influence the practice of safety culture transformation. However, another critical topic that might have been overlooked in the literature is the potential deterioration of the safety culture. In the medical context the safety culture has been found to potentially decline over time from the healthcare workers' perception (Nordin et al., 2020). A similar deterioration of safety culture is also identified in the healthcare workers due to systemic factors (Ling et al., 2016). Nevertheless, several external factors have been revealed to potentially influence an organisation's safety culture at the operational and organisational levels. For instance, introducing new technical systems to air traffic control weakened the perception of workers about safety values (Patankar, 2012). Most deterioration is observed at the operational level (i.e., workers' perceptions) and the primary suggestions provided in the literature are to build a monitoring system and regular assessment (Khan et al., 2010; Kojima et al., 2009; Sanusi et al., 2015), which has been adopted in the nuclear power industry.

On the other hand, another argument is proposed that the improvement process of the safety culture might reach an irreversible and unpredictable condition in the long term. Once significant internal or external changes occur, such as increasing pressure on productivity or introducing new technologies, the stability of the safety culture might relatively deteriorate until another new accident occurs (Berglund, 2020; Younes, 2005). A similar circumstance is also identified in the railway industry and the need to manage the mindset of human factors is also proposed (Tang et al., 2022; Teperi et al., 2023). In addition, an extensive body of literature in railway management focuses on assessing safety culture performance, and human factors dominate the result and discussion (Kalem et al., 2021; Srathongkhruen & Fraszczyk, 2021; Wang & Liu, 2012).

Additionally, the organisational social capital has been proven to weaken safety culture and contribute to accidents (Rao, 2007). Organisational social capital is a collective value of social networks, including characteristics of social organisation such as trust, norms and networks for facilitating cooperation to obtain mutual benefit (Putnam, 2000). Despite the benefit brought by the social capital, excluding external information and overemphasising a strong safety culture without translating into any use of novel technology might result in a conservative and standstill culture (Rao, 2007). Such enforceable trust and faulty value systems have been found to be the primary motivations leading to the deterioration of the safety culture.

In practice, safety is usually managed and controlled by the safety management system (SMS). The aim of a SMS is to create a risk-free environment for workers and the public and represent the arrangement and preparation of safety control for a better safety culture (Kalem et al., 2021). SMS covers a wide range of operational practices and systems depending on the context. For example, an SMS can be an integrated approach for the management of safety across various sectors (ICAO, 2007). SMS can also be any arrangement for actions taken by stakeholders to ensure safety during operations (ERA, 2007). It contains several key aspects directing operators across industries to manage risk systematically and derive many risk management models, such as safety risk management, safety assurance, safety policy and safety promotion (Li & Guldenmund, 2018). Despite a wide discussion of SMS across industries in the literature, most approaches apply accident case-based analysis as the input to produce management insights and prevent historical incidents from occurring (Dai & Wang, 2010; Lin et al., 2020; Rausand, 2013). Cross-jurisdiction learning is notably absent in the SMS framework and might hinder operators from learning safety knowledge from external sources or contexts.

Although the discussion about potential railway safety deterioration is limited in the literature, a number of trends have been observed from the survey of the scoping workshop. For instance, the legislative framework mentioned in Section 5.6.2 has been a significant restriction for practitioners to be conservative in the safety-related decision-making process. This might be considered an organisational social capital because external information and resources have been included in the legislative frameworks, resulting in the declining willingness to introduce novel technologies and higher complexity in processing information proactively. Despite not being a direct indication of deterioration, the difficulties of aggregating historical data might restrict the scale of analysis and limit the perspective of findings. This might potentially lead to a decrease in the retention of prior knowledge and impact long-term railway safety given that knowledge needs to be manually processed while people are constantly changing.

Figure 7-4 illustrates the integrated safety culture framework and the potential relationship between safety culture development and deterioration. Several works have discussed what factors would enable safety behaviours driven by enabling factors. Bisbey et al. (2021) summarised seven critical enabling factors and four enacting behaviours to develop a safety culture and ensure safety outcomes for general industries. Patankar and Sabin (2010) provided approaches for assessing the performance and attitude of the safety culture from the systemic perspective to maintain and monitor an organisation's safety culture. These findings have systematically developed a robust safety culture structure and offer a variety of strategies to the industry for practical implementations. However, several internal and external factors influence the functionality of the industry and might cause unseen hazards. For instance, severe climate change might expose the system to extreme weather conditions that have never occurred. The safety culture perception of frontline workers might also decline over time due to the dynamic social capital mentioned. Therefore, the safety culture might gradually deteriorate until another accident occurs.

On the other hand, there might be several barriers to advancing safety culture while avoiding the deterioration of the safety culture, depending on the nature of the context. For the railway industry, the legislative framework, regulations and accessible supporting technologies for analysing a large amount of data are observed as potential barriers and were discussed in the analysis of scoping workshops and surveys (Section 7.2). Note that these factors cannot represent the comprehensive coverage of barriers but were observed during the analysis in this study.

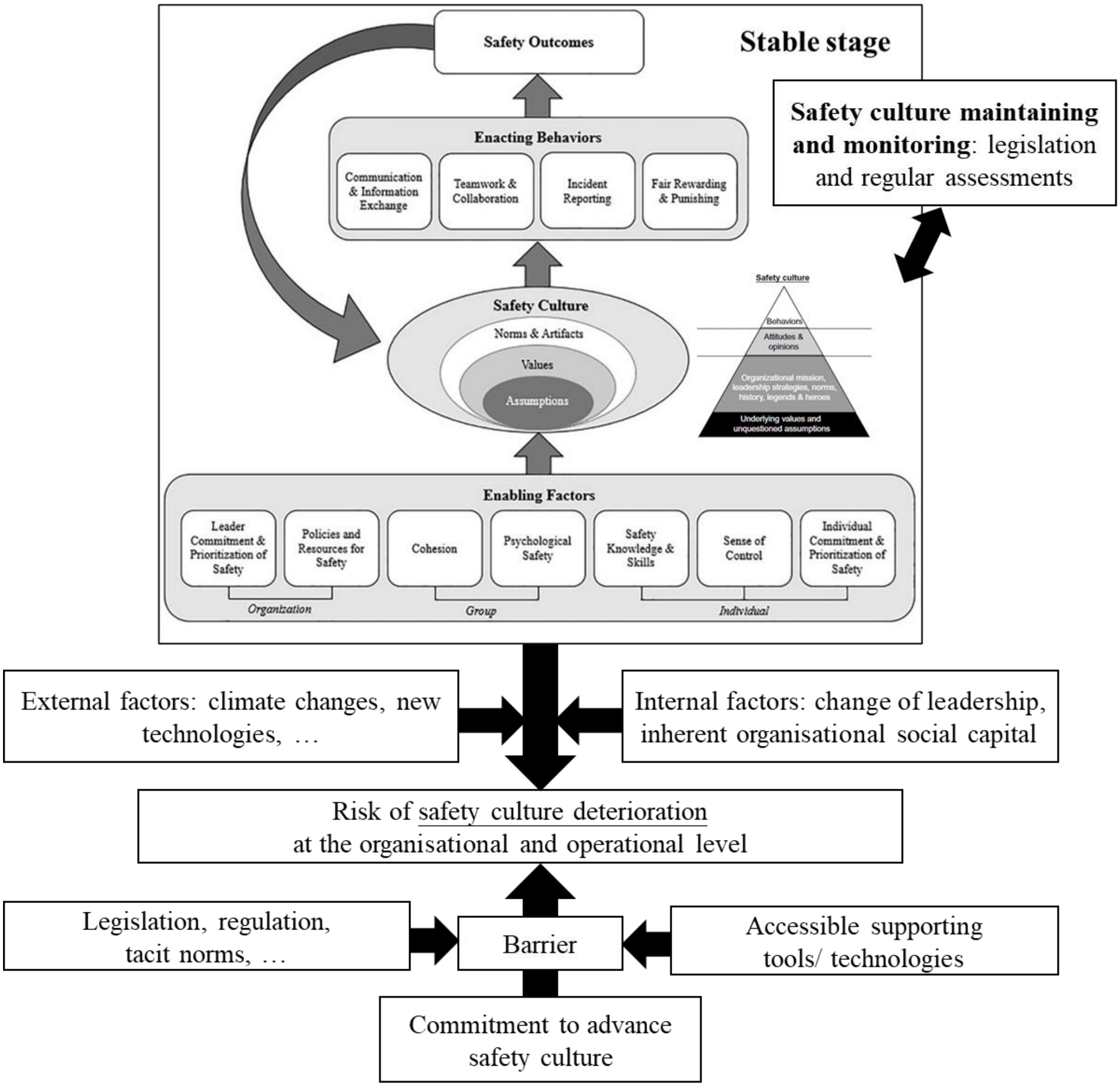


Figure 7-4: Integrated safety culture framework and the potential relationship between safety culture development and deterioration, based on Bisbey et al. (2021) and Patankar and Sabin (2010)

7.4 The opportunity to overcome barriers

The railway industry has devoted much effort to learning over time by accumulating experience and knowledge from previous accidents and constructing a comprehensive knowledge base. However, several barriers to advancing the safety culture discussed above that may potentially lead to the deterioration of the safety culture have been investigated. In addition, external and internal factors might become the catalyst for accelerating such deterioration. For instance, unexpected accidents may occur from a change in the environmental condition such as extreme weather or the implementation of new policies.

Given that *HazardMap* has illustrated that hazards would not be revealed until an accident is triggered, learning from a wider range of resources can be beneficial because various dynamic conditions and hazards might have triggered an accident in other jurisdictions and can be investigated and mitigated in advance. Understanding diverse hazards and aspects can be considered as the “breadth” of understanding of hazards and their combinations assuming that a catalogue containing hazards and aspects is identified across jurisdictions and time. Therefore, applying *HazardMap* to external resources enables users to summarise knowledge of hazards with little human intervention and helps to advance learning behaviours.

Addressing hazards would be the focus after the identification process. In most cases, issues raised by a hazard would cover from the operational to the organisational level and legislative context from the systematic perspective. Therefore, structured approaches for mitigating hazards need to involve a series of contexts, such as risk evaluation, measure development, monitoring and review processes, and dynamic modifications. Adopting multiple dimensions at one time would be extremely difficult without referencing other knowledge and resources. This can be thought of as the “depth” of understanding of a specific hazard. The depth is determined by the extent of the railway industry’s knowledge of how a hazard affects the railway at various organisational levels. To obtain the best of knowledge, *RecoMap* would be a useful tool providing a comprehensive reference for how hazards are handled across jurisdictions and time, which can be considered a baseline for addressing known hazards.

The concept of the “breadth” and “depth” of understanding of hazards can also be utilised to enhance the SMS by bridging the knowledge and experience gap between stakeholders across countries. The cross-jurisdiction learning has not been widely discussed in the SMS context, leaving the integration of organisational learning into operational decision-making remains unclear. The *HazardMap* and *RecoMap* frameworks proposed in this research serve to aggregate, standardise and disseminate incidents and safety knowledge learned. Practitioners in HROs such as the railway industry are enabled to efficiently identify hazards and access knowledge globally to augment the overall safety landscape.

Extending the breadth and depth of understanding to eliminate the deterioration of railway safety culture is a dynamic process. Although the unknown combination of hazards cannot be anticipated, known combinations should have been proactively addressed by reviewing data from multiple sources and reinforcing the railway system. Except for the use of appropriate technologies, constant motivation would be the most critical factor in determining the quality of railway safety. Components related to motivation are complicated and involve multiple dimensions and organisational behaviours. The scoping workshops and surveys have revealed some of the considerations that practitioners are concerned about while handling railway accidents, which might be worth further investigation.

7.5 Synthesis of findings

This chapter further connects the findings of *HazardMap* and *RecoMap* to the learning behaviours in the railway industry. The result of applying *HazardMap* suggests homogenous aspects of a hazard regardless of the locations of investigating bodies and the existence of underlying unseen factors for one jurisdiction that can be addressed in advance by learning from experience in other countries. On the other hand, the result of applying *RecoMap* indicates a growing trend of the safety culture supporting learning behaviours across organisations. However, such learning behaviour has been restricted to local jurisdictions instead of across countries.

From the perspective of learning behaviours in other contexts, many industries have adopted different theories to manage risks and maintain the motivation of learning. For instance, peer review, peer coaching, team behaviour, reporting and systematic analysis have supported the knowledge retrieving, processing and disseminating process in the healthcare system. In contrast, the aviation industry mitigates risks by standardising policies, norms and other latent factors that might influence safety performance as the basis of international aviation operation regulation. For the railway industry, the scoping workshops and surveys have partly revealed concerns hindering practitioners from advancing learning behaviour such as the legislative framework, the lack of a robust knowledge extraction process, and the lack of an international centralised authority managing the knowledge retrieving, processing and disseminating.

Furthermore, potential barriers to learning across jurisdictions and time discussed might deteriorate the organisational safety culture and endanger railway safety. Several internal and external factors that might influence the safety culture development have been discussed and literature relating to the transformation of safety culture at the organisational level has been reviewed. The result suggests that although no direct safety deterioration is found in the railway industry, many signs indicating the potential standstill of safety culture have been observed from *HazardMap* and *RecoMap* and from the scoping workshops and surveys such as organisational social capital and obstacles to introducing new technologies. Nevertheless, being unable to analyse the enormous amount of historical data restricts the scale of knowledge and results in a decline in the retention of prior knowledge. These limitations mentioned above might not cover comprehensive factors causing the deterioration of safety culture but existing evidence and barriers identified support the conclusion that the railway industry may be exposed to the risk of safety deterioration unless further strategies are put in place to stimulate a learning culture.

This chapter summarised the learning behaviours of the railway industry across four jurisdictions by connecting the findings of *HazardMap* and *RecoMap* to existing theories in the literature and comparing them with other industries. Several barriers to learning across jurisdictions and time have been identified to potentially lead to the deterioration of safety culture. External and internal factors influencing the railway safety culture have been discussed as well. Extending the breadth and depth of understanding to eliminate the deterioration of railway safety culture is proposed. The use of *HazardMap* and *RecoMap* is suggested to overcome the difficulties of reviewing data from multiple sources to enhance the railway safety culture and mitigate the risk of safety culture deterioration.

8. Conclusions

This chapter summarises findings, highlights the contributions and limitations of this thesis and makes suggestions for future work. The structure of the chapter is as follows: the key findings of this thesis are presented (Section 8.1), followed by highlighted contributions (Section 8.2), limitations (Section 8.3), and future work and challenges (Section 8.4).

8.1 Key findings

This section discusses and highlights key findings from the literature review (Chapter 2), methodological literature review (Chapter 3), methods for developing the models *HazardMap* (Section 6.3) and *RecoMap* (Section 6.2), and results of analysis supplemented with evidence in the literature.

8.1.1 Key findings from the literature review

First, conflicts in the literature in defining vulnerability assessment due to divergent points of view were revealed, and definitions of vulnerability and vulnerability assessment in the context of the transport system were reviewed. Most studies on vulnerability assessment in the transport system are based on reliability theory, probability theory, and statistical and optimisation approaches.

Second, the review also revealed that the complexity of socio-technical systems has led to the development of frameworks like software–hardware–environment–liveware (SHELL) and human factors analysis and classification system (HFACS), but data collection and analysis remain challenging. Interface weakness is also identified as a factor in railway accidents, but its importance is not fully understood in literature due to a lack of understanding of the causes and sequences of railway incidents and the application of historical accident textual data.

Last, the review suggested that natural language processing techniques have strong potential to mitigate the gaps identified in railway safety by extracting critical hazards from accident records although such technology is still partly reliant on human determination, making the process time-consuming. Some studies have proven such potential by building an ontology to represent domain knowledge and combining it with natural language processing techniques to improve hazard identification (Debbech et al., 2020; Hulin et al., 2016; Zhao et al., 2022). However, the implementation of natural language processing to real-world data for hazard mitigation strategies remains a significant challenge.

8.1.2 Key findings from the methodological literature review

The review of methodological literature provided an in-depth introduction to natural language processing models and training approaches, including word embedding and state-of-the-art language models. The benefits and drawbacks of different training approaches and the concept of ontology in the context of railway accidents were discussed. The limitations of existing railway-related ontologies and the challenges of reusability were highlighted. Knowledge graphs and the entity linking process, emphasising their application and evaluation, were introduced. Findings indicated that novel techniques using off-the-shelf tools have strong potential to overcome the limitations of overreliance on manual analysis in practice and theory, but the absence of shared railway safety-related benchmark corpora restricts implementation.

8.1.3 Key findings from methods for developing models

Several analysis approaches were reviewed to develop models to address the research questions (Section 8.3). First, several analysis functions required are illustrated, including topic modelling, entity linking strategy, covariate analysis, temporal analysis and model evaluation. Next, a comparison was made between potential candidates for topic modelling, considering data dimensionality, analysis toolkits, and outcomes. Due to their respective strengths, BERTopic was selected to analyse accident reports at the sentence level, while STM was chosen to analyse accident recommendations. Subsequently, the entity linking strategy was illustrated, involving the establishment of an initial railway accident ontology and the use of knowledge graphs to identify and disambiguate entities. Wikidata was selected as the appropriate knowledge graph, and the *Tagme* API toolkit was used to supplement the ontology. Context-sensitive disambiguation based on graph theory was proposed to enhance the model's effectiveness. The developed railway accident ontology was linked to real-world data, primarily for standardising terminology and bridging the gap between topic modelling keywords and original accident reports. Last, scoping workshops and a survey were conducted to evaluate the developed model, gathering feedback and observations from the railway industry to assess its consistency and potential application in practice.

8.1.4 Key findings from HazardMap

HazardMap proposed in this thesis was used to analyse over 1,200 railway accident reports from four countries (the UK, US, Australia, and Canada), with a total of 400,000 sentences, to identify specific hazards and understand the differences in accident mechanisms and approaches to addressing them. To evaluate the performance of models, the coefficient of variance (CV) indicator was designed to effectively discriminate distinguished topics from common ones, enabling the identification of specific hazards. *HazardMap* is also capable of visualising the nature of hazards, their mechanisms, and the relationships between hazards, accidents and recommendations across countries. On the other hand, *HazardMap* suggests that a hazard must have multiple aspects that result in different types of accidents. Therefore, an addressed hazard might appear again after combining with others, implying one hazard may never be fully addressed. However, reducing the caused impact by proposing appropriate recommendations for accidents revealing part of the aspects of one hazard is still beneficial for improving railway safety. Additionally, it also reveals that each country had different dominant hazards, such as tram-related accidents being a major hazard in the RAIB dataset from the UK and freight train-related accidents being prevalent in the ATSB dataset from Australia. Last, *HazardMap,* with the support of the ontology developed, offers a practical approach to incorporate data retrieved from external sources, providing a consistent and standardised framework for practitioners to process and archive the railway accident knowledge acquired from the investigation with limited human intervention required.

8.1.5 Key findings from RecoMap

RecoMap developed by this thesis observed that different investigating bodies had varying styles of making recommendations, focusing on improving systems, taking action on safety issues, cooperating with other organisations, or ensuring effective procedures and designs. It also revealed some limitations in the recommendations provided. Recommendations often focused on introducing modifications or monitoring systems without discussing underlying motivations or conducting day-to-day analysis. Additionally, reviewing compliance with procedures or regulations implied a blame culture, contrary to the promises made by investigators. Finally, *RecoMap* was proposed as a solution to overcome limitations in understanding the perspective of each recommendation and the comprehensive overview of recommendations made by investigators. *RecoMap* helped reveal recommendation styles, distinguish the role each recommendation plays in the railway industry, and provided insights for authorities and investigators to concentrate on appropriate solutions based on the jurisdiction, promote engagement and knowledge sharing across jurisdictions and enable comprehensive views from the perspective of the socio-technical hierarchy in the railway industry.

8.1.6 Key findings from the results of analysis supplemented with evidence in the literature

Several key findings of the analysis were extracted and discussed. First, *HazardMap* revealed homogenous aspects of hazards regardless of jurisdiction and the existence of unseen factors that can be addressed by learning from other countries. *RecoMap* showed a growing trend of safety culture supporting learning behaviours within organisations but limited to local jurisdictions. On the other hand, the scoping workshops and survey highlighted concerns like the legislative framework, lack of robust knowledge extraction, and absence of a globally centralised authority for knowledge management. Therefore, the analysis indicated potential barriers to learning across jurisdictions and time may weaken safety culture. Factors like organisational social capital and obstacles to adopting new technologies indicate a potential standstill in safety culture. To overcome the limitations identified, analysing historical data and implementing *HazardMap* and *RecoMap* may enhance railway safety culture and mitigate the risk of deterioration in the safety culture.

8.2 Highlighted contributions

This thesis presents several insights and advancements in the context of railway safety across countries, railway vulnerability analysis, implementation of natural language processing, ontology and knowledge graphs, learning behaviours in the railway industry and railway safety culture. The following key contributions have been achieved through rigorous research, analysis and discussion:

(1.) Contributions to reviews of the literature body of railway systems vulnerability analysis and methodology of implementation: This thesis provided a comprehensive review of the literature body in the field of vulnerability analysis and methodology used in risk assessment. The review summarised past progress, current trends and future directions and discussed possibilities, challenges and limitations.

(2.) Development of a generalised and data-driven framework for understanding hazards: This thesis proposed the *HazardMap* framework to depict the nature of hazards in the railway system and their mechanisms illustrated by a case study comparing four different countries. Over 1,200 railway accident reports, containing 400,000 sentences, published by national railway accident investigation bodies of four countries were analysed. This enabled a huge amount of knowledge accumulated for an intuitive policymaking process to be summarised and allows other railway investigators to leverage lessons learned across jurisdictions and time with limited human intervention.

(3.) Providing a systematic view of recommendations made by investigators: This thesis analysed over 9,000 sentences in the recommendation section of railway accident reports published by RAIB, ATSB, NTSB and TSB. The structural topic model has been applied to explore latent topics within each dataset, enabling us to understand the emphasis investigators put on mitigating hazards identified. The developed model *RecoMap* is proposed to describe the distribution of recommendations made from the organisational perspective across different countries, providing an alternative approach for interpreting the outcomes of topic modelling, which prior works have struggled with. It also allows the railway industry to learn across jurisdictions and time by offering a systematic view of recommendations made in different jurisdictions. Several limitations remain and are worth investigating (Section 8.4).

(4.) Development of a novel framework for interpreting outcomes of an unsupervised learning-based natural language processing model: This thesis created a novel framework for the systematic interpretation of outcomes of an unsupervised learning-based natural language processing model by leveraging the application of ontology and knowledge graphs for standardising terminology used by different regions, revealing the relationship between topics extracted and designing advanced analysis applications. Additionally, all tools used in the framework are off-the-shelf and open-access online. This availability is a valuable asset for researchers and practitioners, enabling them to conduct further studies, develop novel algorithms, and evaluate existing approaches in a standardised manner.

(5.) Revealing potential trends in the way that railway accident investigations make recommendations: This thesis identified a transition from making interfering recommendations addressing operational issues to making supportive recommendations addressing organisational issues in the railway industry across countries. A growing trend of promoting learning across jurisdictions and knowledge sharing was also found across the investigating bodies. These findings provide valuable insights for managers and practitioners on how to systematically react to railway accidents through effective decision-making practices tailored to specific cultural environments.

(6.) Understandings of learning behaviours in the railway industry and railway safety knowledge retrieving, processing and disseminating: This thesis integrated theoretical perspectives such as the resource-based view and inter-firm network theory to analyse knowledge flow and its impact on the formation of safety culture in the railway industry and discussed the importance of information flow and knowledge extraction processes in the railway industry, highlighting concerns about legislative frameworks, the cost and expected performance of safety improvements, and the lack of adequate technology for analysis. A collection of insights into the gaps, challenges and potential strategies for promoting organisational learning and safety culture in the railway industry were provided, enabling researchers and practitioners to understand the barriers and incentives that influence behaviour of learning across jurisdictions and time.

(7.) Revealing the potential deterioration in safety in the railway industry due to the limitation of learning behaviours: This thesis highlighted the overlooked topic of the potential deterioration of safety culture over time and provided examples from the healthcare and aviation industries, emphasising the need for monitoring systems and regular assessments to prevent decline. The analysis also suggested that the improvement process of safety culture may reach an irreversible and unpredictable condition in the long term. Internal and external changes, such as increased pressure on productivity or the introduction of new technologies, can lead to a deterioration of the safety culture until another accident occurs. These findings provide a unique insight for policymakers and practitioners to recognise an emerging hazard that has not been fully identified.

8.3 Outcome of Research Questions

This section discusses the response to each research question with findings and evidence. Table 8-1 summarises the results for the research questions and corresponding propositions.

Table 8-1: The result for research questions and corresponding propositions

Research questions	Proposition	Result
RQ1-1: What is the difference in roles each hazard plays in various jurisdictions during railway accidents?	Regardless of hazard taxonomy, certain hazards have similar attributes and have resulted in comparable railway accidents across different jurisdictions.	● Difficulty in efficiently identifying hazards is due to the analysis of large amounts of safety-related textual data ● *HazardMap* demonstrates comparable hazards across jurisdictions, triggering similar accidents ● Hazards have been identified to have impacts in multiple countries and can trigger comparable railway accidents
RQ1-2: Do the same hazards occur in different jurisdictions and across time?	There are some hazards sharing similar characteristics and occurring in different jurisdictions and across time.	
RQ1-3: Do those hazards result in similar vulnerabilities in different jurisdictions and times?	Those hazards sharing similar characteristics and occurring in different jurisdictions and across time may result in similar vulnerabilities and railway accidents.	● Similar hazards may trigger comparable accidents across jurisdictions with similar mechanisms ● Hazards have resulted in similar vulnerabilities regardless of accident severity
RQ2-1: How do recommendations made by railway accident investigators address hazards identified from the socio-technical perspective?	Recommendations made by railway accident investigators might primarily focus on addressing operational issues and concentrate less on risks at the management level.	● The purpose of recommendations in the railway industry is to improve safety, not assign blame ● It is difficult for the railway industry to determine the scope of and understand previous recommendations ● The majority of recommendations address operational issues, but a growing trend focuses on addressing organisational issues ● There is a transition from interfering recommendations to supportive recommendations in the railway industry
RQ2-2: Is there a transition in the style of making railway accident recommendations in each jurisdiction over time?	The style of making railway accident recommendations might change over time, resulting in a potential change in the way that recommendations are proposed.	

Table 8-1: The result for research questions and corresponding propositions (continued)

Research questions	Proposition	Result
RQ3-1: Do railway accident report recommendations support the railway industry to learn across jurisdictions and time?	Recommendations made in railway accident reports often overlook the importance of learning across jurisdictions and time.	● Limited information exchange and co-reference between railway jurisdictions is observed ● Limited recommendations published for promoting learning behaviours are found ● Insufficient implementation of learning across time in the railway industry is recognised
RQ3-2: What are potential barriers to the railway industry learning across jurisdictions and time?	The railway industry has multiple barriers to learning across jurisdictions and time in the socio-technical hierarchy.	● Issues such as language barriers and legal obligations hinder the efficiency of learning ● Restrictions of regulations and policies, authority-centralised information sharing culture, and organisational social capital act as barriers
RQ3-3: What hazard(s) might emerge if barriers to learning across jurisdictions and time remained unsolved?	The absence of learning across jurisdictions and time might significantly impact safety culture and lead to a deterioration.	● Decreased retention of prior knowledge and impact on long-term railway safety are identified ● Risk of safety culture deterioration at organisational and operational levels due to the absence of learning across jurisdictions and time is observed.

For RQ1-1 and RQ1-2, the results of the literature review reveal that several theories are used as the foundational basis to build a model for railway hazard analysis and accident prevention strategies to comprehensively interpret the epidemiology of accidents in modern society (Section 2.2.1), indicating the attempts of the literature to generalise the nature of hazards regardless of investigated locations. However, hazards cannot be identified efficiently due to the inability to analyse a large amount of safety-related textual data (Section 2.4 and Hong et al., 2023). On the other hand, the *HazardMap* developed (Section 6.3) has shown that hazards are potentially comparable across jurisdictions and aspects of these hazard did trigger similar accidents. For instance, factors such as sighting distance and audibility have shown impacts in more than one country and these factors are capable of triggering comparable railway accidents at level crossings (see Section 5.8). Therefore, it is argued that similar hazards may occur in different jurisdictions and across time with similar roles during railway accidents.

For RQ1-3, the discussion of how hazards in the railway system result in vulnerabilities was limited (Section 2.2 and Hong et al., 2022). However, the case study of level crossing accidents (Section 6.5) and the platform–train interface risk (Section 6.6) and the cross-sectional analysis (Section 5.5.2) demonstrated that similar hazards triggered railway accidents across jurisdictions through a variety of aspects and their interactions (Section 6.3), although the extent to which these hazards make the railway system unsafe is not analysed. Therefore, these hazards have led to similar vulnerabilities regardless of the severity of accidents.

For RQ2-1 and RQ2-2, some evidence from practitioners has shown that the purpose of recommendations made in the railway industry is not for blame but to help railway operators or regulators improve safety (Section 2.3.1). However, discussion about the role of recommendations is rare in the literature. Preliminary findings of previous studies only indicate that investigating bodies have difficulty in determining the scope of recommendations and systematically learning and understanding recommendations made in previous railway accidents (Section 2.3.3). On the other hand, *RecoMap* indicates, as discussed in Section 6.2, that the majority of recommendations made by railway accident investigators aim to address issues at the operational level. However, a growing trend to mitigate risks from organisational levels is also identified in *RecoMap*. For example, recommendations made by RAIB and NTSB after 2010 focus more on addressing organisational issues, such as standardising processes and developing procedures systematically (see Figure 6-2 in Section 6.2). This might imply a transition from making interfering recommendations addressing operational issues to making supportive recommendations addressing organisational issues in the railway industry across countries.

For RQ3-1, the importance of learning behaviours has been widely discussed in many contexts (Section 7.2). Current learning behaviours in the railway industry were also discussed (Section 7.1). Co-reference between RAIB, NTSB, TSB and ATSB as discussed in Section 5.5.2 showed that information exchange across jurisdictions in the railway industry is limited. On the other hand, the scoping workshops and survey showed that the railway industry has been divided into many organisations in various disciplines functioning individually and most legislative systems in jurisdictions assign the obligation of responding to recommendations made by investigators to railway organisations. In other words, investigating bodies play the major role in promoting learning behaviours. However, *RecoMap* suggests that recommendations supporting learning behaviours are limited (Section 6.2). The scoping workshops and survey also indicated that learning across time has been implemented in the railway industry, which has also been supported by the literature and this thesis (Sections 5.5.2 and 6.1).

For RQ3-2, several barriers have been identified by the scoping workshops and survey and the interpretation of findings from this research supplemented by additional evidence. For example, the scoping workshops and survey revealed that despite the increased awareness of the importance of how to efficiently implement recommendations, the industry might have ignored the issue of the efficiency of learning across jurisdictions due to several issues, such as language barriers and legal obligations (Section 7.2). On the other hand, the interpretation of findings shows that restrictions of regulations and policies, the culture of an authority-centralised information sharing system and the inherent organisational social capital might result in potential barriers to the railway industry learning across jurisdictions and time (Sections 7.2 and 7.3).

For RQ3-3, the interpretation of findings indicated that the difficulties of aggregating historical data might restrict the scale of analysis and limit the perspective of findings. This might potentially lead to a decrease in the retention of prior knowledge and have an impact on long-term railway safety given that knowledge needs to be manually processed while people are constantly changing (Section 7.3). Additionally, several internal and external factors influence the functionality of the industry and might cause unseen hazards. For instance, severe climate change might expose the system to extreme weather conditions that have never occurred in that location. The safety culture perception of frontline workers might also decline over time due to the dynamic social capital. Therefore, the safety culture might gradually deteriorate until another accident occurs (Section 7.3), resulting in the risk of safety culture deterioration at organisational and operational levels as the result of the absence of learning across jurisdictions and time.

8.4 Limitations and further research

Despite the contributions discussed, several primary limitations in this thesis are identified:

(1.) For the performance of models, this thesis restricted the countries analysed based on several factors such as using English language, the level of development of railway safety systems and the number of accident reports published. This limitation might overlook non-English speaking countries with a fully developed railway safety industry such as Japan. Future research could overcome such a limitation by introducing more advanced NLP technology such as Large Language Models (LLMs).

(2.) Only off-the-shelf programming packages were used in building models, which might lead to limited flexibility to customise the functions needed. The effect of each factor extracted on the consequences of railway accidents cannot be fully discovered by the models presented due to the absence of combining with numerical data analysis. This might make railway safety policymakers unable to prioritise strategies for addressing hazards. Future research could consider integrating NLP models with other statistical models to further investigate the importance of each hazard identified.

(3.) The outcome of *HazardMap* relies heavily on the characteristics of the input data, meaning that critical features missing in the original data would result in the absence of features in the constructed *HazardMap*. Human interpretation might still be required while processing systematic factors or underlying causes. The name of each extracted topic also needs to be determined manually by reviewing keywords of each topic.

(4.) *RecoMap* has not yet revealed the incentives and barriers making the railway industry follow safety-related instructions that are critical for decision-makers to understand the behaviour of practitioners. Future works could consider using advanced qualitative approaches such as grounded theory or ethnographic studies to delve deeper into the professional experiences and cultural norms.

(5.) The performance and effectiveness of recommendations made by different countries are difficult to evaluate and compare in *RecoMap* although they have played an important role in cost-benefit analysis. The interpretation of topics extracted from the topic model is time-consuming and still requires manual effort. Future research is suggested to leverage LLMs to support the topic interpretation to increase the efficiency, ensure consistency and reduce human error.

(6.) A limited number of respondents participated in the scoping workshops and survey due to the voluntary nature of recruitment and a lack of external incentives. More experts from various jurisdictions, especially non-English speaking countries, should be included for a diverse perspective. Future work might expand the number of countries investigated for a more comprehensive analysis.

Future work might consider addressing the limitations above and concentrate on potential solutions for reducing human intervention required while interpreting results from the topic model. A shared decision-making platform based on *HazardMap* might also be worth investigating. The connection between hazards and aspects in *HazardMap* and recommendations in *RecoMap* also requires more work to identify the relationship between them and evaluate the effectiveness of recommendations by introducing data on the consequences of accidents. On the other hand, more understanding of deterioration of safety culture in the railway industry needs to be investigated from the perspective of the railway industry across jurisdictions. More evidence is required to reveal the dynamics of safety culture deterioration in the railway industry. Last, further work is also encouraged to apply

HazardMap and *RecoMap* to incorporate data from other contexts, such as road, aviation or maritime accidents.

9. Appendix

9.1 Appendix A. the outlines of the participants' survey

The survey consists of the following topics:

(1.) **Background information:** organisations, job title, experience, and main responsibilities, etc.

(2.) **Information receiving:** open-ended questions primarily concentrating on the current approach for practitioners to collect data, including: what type of information / evidence do you think could help in decision-making in terms of promoting rail safety? Where does your organisation most frequently source the information or instructions relating to rail safety? Which type of information relating to rail safety draw your organisation's attention the most? and which are the most important sources?

(3.) **Information processing:** open-ended questions discussing about how railway practitioners process and analyse data or information collected, including: how does your organisation retrieve and process the historical accident reports? Does your organisation also consider the reports from other countries or jurisdiction? What makes your organisation not willing to consider using historical accident reports? Please specify. Did you ever go through the partial or whole railway accident report before making any railway safety-related decision? how you used a railway accident report to help in your decision-making, Which parts of that railway accident report do you think could support your decisions? Please briefly explain the reason. What method/ evidence/ knowledge you currently (or previously) use to support your decisions? What railway safety-related model(s) you are currently using in the decision-making? And what makes you select the model(s)?

(4.) **Information disseminating:** open-ended questions reflecting the current approaches and obstacles and concerns for practitioners to sharing knowledge, including: when a major railway accident occurs and your organisation is involved, would your organisation have any engagement relating to investigation or improvement with others? Please specify the organisation you engage with the most and briefly describe the process of engagement, please briefly describe the difficulties your organisation faces when engaging with others, would your organisation report / share any information or knowledge after major conclusions or learning points have been in place?

9.2 Appendix B. the approval letter from Human Research Ethics Committee at the University of Sydney

Research Integrity & Ethics Administration
HUMAN RESEARCH ETHICS COMMITTEE

Monday, 18 July 2022

Dr Geoffrey Clifton
Institute of Transport and Logistics Studies (ITLS); University of Sydney Business School
Email: geoffrey.clifton@sydney.edu.au

Dear Geoffrey,

The University of Sydney Human Research Ethics Committee (HREC) has considered your application.

I am pleased to inform you that after consideration of your response, your project has been approved.

Details of the approval are as follows:

Project No.:	**2022/268**
Project Title:	**Scoping workshop survey: Utilising Natural Language Processing to enhance rail safety, learn from the history and improve the dissemination of knowledge**
Authorised Personnel:	**Clifton Geoffrey; Hong Wei-ting; Nelson John;**
Approval Period:	**18/07/2022 to 18/07/2026**
First Annual Report Due:	**18/07/2023**

Documents Approved:

Date Uploaded	Version Number	Document Name
01/07/2022	Version 2	Participant Consent Form_Final
01/07/2022	Version 2	Participant Information Statement_Final

For Noting

- Please ensure the version numbers and date in the footer of all public documents are updated and revise these any time changes are made to the documents. This allows identification of original, updated, and approved versions in IRMA and to facilitate any future requested modifications.

Condition/s of Approval

- Research must be conducted according to the approved proposal.
- An annual progress report must be submitted to the Ethics Office on or before the anniversary of approval and on completion of the project.
- You must report as soon as practicable anything that might warrant review of ethical approval of the project including:
  - Serious or unexpected adverse events (which should be reported within 72 hours).
  - Unforeseen events that might affect continued ethical acceptability of the project.
- Any changes to the proposal must be approved prior to their implementation (except where an amendment is undertaken to eliminate *immediate* risk to participants).
- Personnel working on this project must be sufficiently qualified by education, training and experience for their role, or adequately supervised. Changes to personnel must be reported and approved.
- Personnel must disclose any actual or potential conflicts of interest, including any financial or other interest or affiliation, as relevant to this project.

Research Integrity & Ethics Administration
Research Portfolio
Level 3, F23 Administration Building
The University of Sydney
NSW 2006 Australia

T +61 2 9036 9161
E human.ethics@sydney.edu.au
W sydney.edu.au/ethics

ABN 15 211 513 464
CRICOS 00026A

10. References

Abdullah, D., Takahashi, H., &Lakhani, U. (2017). Improving the Understanding between Control Tower Operator and Pilot Using Semantic Techniques - A New Approach. *Proceedings - 2017 IEEE 13th International Symposium on Autonomous Decentralized Systems, ISADS 2017*, 207–211. https://doi.org/10.1109/ISADS.2017.8

Adeolu, O. D., Cornelius, O. A., &Bamidele, A. B. (2016). Evaluation of railway level crossing attributes on accident causation in Lagos, Nigeria. *Indonesian Journal of Geography*, *48*(2), 108–117. https://doi.org/10.22146/ijg.17520

Aho, A., Sethi, R., &Ullman, J. (1963). *Compilers: Principles, Techniques, Tools.* MA: Addison-Wesley.

Akel, A. J. N., Gravio, G.Di, Fedele, L., &Patriarca, R. (2022). Learning from Incidents in Socio-Technical Systems: A Systems-Theoretic Analysis in the Railway Sector. *Infrastructures 2022, Vol. 7, Page 90*, *7*(7), 90. https://doi.org/10.3390/INFRASTRUCTURES7070090

Alashri, S., Tsai, J. Y., Koppela, A. R., &Davulcu, H. (2018). Snowball: Extracting causal chains from climate change text corpora. *Proceedings - 2018 1st International Conference on Data Intelligence and Security, ICDIS 2018*, 234–241. https://doi.org/10.1109/ICDIS.2018.00045

Alexakis, D. D., Agapiou, A., Tzouvaras, M., Themistocleous, K., Neocleous, K., Michaelides, S., &Hadjimitsis, D. G. (2014). Integrated use of GIS and remote sensing for monitoring landslides in transportation pavements: The case study of Paphos area in Cyprus. *Natural Hazards*, *72*(1), 119–141. https://doi.org/10.1007/s11069-013-0770-3

Alpizar-Chacon, I., &Sosnovsky, S. (2019). Expanding the Web of Knowledge: One Textbook at a Time. *Proceedings of the 30th ACM Conference on Hypertext and Social Media*, 10. https://doi.org/10.1145/3342220

Alt, C., Hübner, M., &Hennig, L. (2019). Fine-tuning Pre-Trained Transformer Language Models to Distantly Supervised Relation Extraction. *ACL 2019 - 57th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference*, 1388–1398. http://arxiv.org/abs/1906.08646

An, M., Chen, Y., &Baker, C. J. (2011). A fuzzy reasoning and fuzzy-analytical hierarchy process based approach to the process of railway risk information: A railway risk management system. *Information Sciences*, *181*(18), 3946–3966. https://doi.org/10.1016/j.ins.2011.04.051

An, N., Xiao, Y., Yuan, J., Yang, J., &Alterovitz, G. (2019). Extracting causal relations from the literature with word vector mapping. *Computers in Biology and Medicine*, *115*, 103524. https://doi.org/10.1016/J.COMPBIOMED.2019.103524

Angelov, D. (2020). *Top2Vec: Distributed Representations of Topics*. https://doi.org/10.48550/arxiv.2008.09470

Apostolakis, G. (1990). The concept of probability in safety assessments of technological systems. In *Science* (Vol. 250, Issue 4986, pp. 1359–1364). American Association for the Advancement of Science. https://doi.org/10.1126/science.2255906

Arenius, M., &Sträter, O. (2014). Resilience engineering in railways-results from a systemic accident and event analysis in German railways. *Safety, Reliability and Risk Analysis: Beyond the Horizon - Proceedings of the European Safety and Reliability Conference, ESREL 2013*, 423–428. https://www.scopus.com/inward/record.uri?eid=2-s2.0-84899995697&partnerID=40&md5=14b941368beba5d99620ea271931e223

Arsene, O., Dumitrache, I., &Mihu, I. (2011). Medicine expert system dynamic Bayesian Network and ontology based. *Expert Systems with Applications*, *38*(12), 15253–15261. https://doi.org/10.1016/j.eswa.2011.05.074

Asim, M. N., Wasim, M., Khan, M. U. G., Mahmood, W., &Abbasi, H. M. (2018). A survey of ontology learning techniques and applications. *Database*, *2018*(2018). https://doi.org/10.1093/database/bay101

ATSB. (2009). *Railway Accident Investigation Guidelines for Railway Network Owners, Railway Operators and Emergency Services Personnel* (3rd ed.). Australian Government.

ATSB. (2011). *Safety issues and safety actions identified through ATSB transport safety investigations: 2010-2011 financial year*.

ATSB. (2019). *Level crossing collision between truck and passenger train 8753, Phalps Road, Larpent, Victoria, on 13 July 2016*.

ATSB. (2020). *Derailment of grain train 8838N, Narwonah, New South Wales, on 1 October 2017*. https://www.atsb.gov.au/publications/investigation_reports/2017/rair/ro-2017-014/

Augenstein, I., Maynard, D., &Ciravegna, F. (2016). Distantly supervised Web relation extraction for knowledge base population. *Semantic Web*, *7*(4), 335–349. https://doi.org/10.3233/SW-150180

Baader, F., Calvanese, D., McGuinness, D., Patel-Schneider, P., &Nardi, D. (2003). *The description logic handbook: Theory, implementation and applications.* Cambridge university press.

Bahdanau, D., Cho, K. H., &Bengio, Y. (2015, September 1). Neural machine translation by jointly learning to align and translate. *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*. https://arxiv.org/abs/1409.0473v7

Balali, A., Asadpour, M., &Jafari, S. H. (2021). *COfEE: A Comprehensive Ontology for Event Extraction from text, with an online annotation tool*. https://arxiv.org/abs/2107.10326v2

Barlow, R. E., &Proschan, F. (1975). *Statistical theory of reliability and life testing: probability models.* Florida State Univ Tallahassee.

Barlow, R. E., &Proschan, F. (1996). *Mathematical theory of reliability*. Society for Industrial and Applied Mathematics.

Bates, S., Angeon, V., &Ainouche, A. (2014). The pentagon of vulnerability and resilience: A methodological proposal in development economics by using graph theory. *Economic Modelling*, *42*, 445–453. https://doi.org/10.1016/j.econmod.2014.07.027

Bayat, S., Cuggia, M., Rossille, D., Kessler, M., &Frimat, L. (2009). Comparison of Bayesian network and Decision Tree methods for predicting access to the renal transplant waiting list. *Studies in Health Technology and Informatics*, *150*(February), 600–604. https://doi.org/10.3233/978-1-60750-044-5-600

Baysari, M. T., McIntosh, A. S., &Wilson, J. R. (2008). Understanding the human factors contribution to railway accidents and incidents in Australia. *Accident Analysis and Prevention*, *40*(5), 1750–1757. https://doi.org/10.1016/j.aap.2008.06.013

Bazovsky, I. (1961). *Reliability Theory and Practice* (Prentice-H).

Beamer, B., Bhat, S., Chee, B., Fister, A., Rozovskaya, A., &Girju, R. (2007). UIUC: A Knowledge-rich Approach to Identifying Semantic Relations between Nominals.

Proceedings of the Fourth International Workshop on Semantic Evaluations (SemEval-2007), 386–389. http://oak.colorado.edu/assert/

Belkacem, Thiziri, Dkaki, Taoufiq, Moreno, José G, Boughanem, &Mohand. (2017). *Apprentissage de représentations de documents et leur exploitation en recherche d'information*. https://oatao.univ-toulouse.fr/22268

Bell, M., &Iida, Y. (1997). *Transportation network analysis*. https://trid.trb.org/view/573501

Benoit, K., Watanabe, K., Wang, H., Nulty, P., Obeng, A., Müller, S., &Matsuo, A. (2018). quanteda: An R package for the quantitative analysis of textual data. *Journal of Open Source Software*, *3*(30), 774. https://doi.org/10.21105/JOSS.00774

Berche, B., VonFerber, C., Holovatch, T., &Holovatch, Y. (2009). Resilience of public transport networks against attacks. *European Physical Journal B*, *71*(1), 125–137. https://doi.org/10.1140/epjb/e2009-00291-3

Berdica, K. (2002). An introduction to road vulnerability: What has been done, is done and should be done. *Transport Policy*, *9*(2), 117–127. https://doi.org/10.1016/S0967-070X(02)00011-2

Berdica, K., &Mattsson, L. G. (2007). Vulnerability: A model-based case study of the road network in Stockholm. In *Advances in Spatial Science* (Issue 9783540680550, pp. 81–106). Springer International Publishing. https://doi.org/10.1007/978-3-540-68056-7_5

Berglund, J. (2020). After Fukushima: Safety culture and fostering critical thinking. *Safety Science*, *124*, 104613. https://doi.org/10.1016/J.SSCI.2020.104613

Berle, Ø., Asbjørnslett, B. E., &Rice, J. B. (2011). Formal Vulnerability Assessment of a maritime transportation system. *Reliability Engineering and System Safety*, *96*(6), 696–705. https://doi.org/10.1016/j.ress.2010.12.011

Bešinović, N. (2020). Resilience in railway transport systems: a literature review and research agenda. *Transport Reviews*, *40*(4), 457–478. https://doi.org/10.1080/01441647.2020.1728419

Binti Sa'adin, S. L., Kaewunruen, S., &Jaroszweski, D. (2016). Heavy rainfall and flood vulnerability of Singapore-Malaysia high speed rail system. *Australian Journal of Civil Engineering*, *14*(2), 123–131. https://doi.org/10.1080/14488353.2017.1336895

Bird, J. F. E., &Germain, G. L. (1986). *Practical loss control leadership.* Loganville, Ga.

Bird, S., &Loper, E. (2004). NLTK: The Natural Language Toolkit. In *Proceedings of the ACL Interactive Poster and Demonstration Sessions* (pp. 214–217). https://aclanthology.org/P04-3031

Bisbey, T. M., Kilcullen, M. P., Thomas, E. J., Ottosen, M. J., Tsao, K. J., &Salas, E. (2021). Safety Culture: An Integration of Existing Models and a Framework for Understanding Its Development. *Human Factors*, *63*(1), 88–110. https://doi.org/10.1177/0018720819868878/ASSET/IMAGES/LARGE/10.1177_0018720819868878-FIG2.JPEG

Bischof, S., &Schenner, G. (2021). Rail Topology Ontology: A Rail Infrastructure Base Ontology. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *12922 LNCS*, 597–612. https://doi.org/10.1007/978-3-030-88361-4_35

Blaho, P., Peceny, L., &Gasparik, J. (2020). Causality of accidents at railway-crossings in Slovakia and its prevention. *2020 12th International Science-Technical Conference AUTOMOTIVE SAFETY, AUTOMOTIVE SAFETY 2020*. https://doi.org/10.1109/AUTOMOTIVESAFETY47494.2020.9293528

Blei, D. M. (2012). Probabilistic topic models. *Communications of the ACM*, *55*(4), 77–84. https://doi.org/10.1145/2133806.2133826

Blei, D. M., &Mcauliffe, J. D. (2007). Supervised Topic Models. *Advances in Neural Information Processing Systems*, *20*. www.digg.com

Blei, D. M., Ng, A. Y., &Edu, J. B. (2003). Latent Dirichlet Allocation Michael I. Jordan. In *Journal of Machine Learning Research* (Vol. 3).

Bonacich, P. (2007). Some unique properties of eigenvector centrality. *Social Networks*, *29*(4), 555–564. https://doi.org/10.1016/J.SOCNET.2007.04.002

Boorjian, S. A., Eastham, J. A., Graefen, M., Guillonneau, B., Karnes, R. J., Moul, J. W., Schaeffer, E. M., Stief, C., &Zorn, K. C. (2012). A critical analysis of the long-term impact of radical prostatectomy on cancer control and function outcomes. In *European Urology* (Vol. 61, Issue 4, pp. 664–675). Elsevier. https://doi.org/10.1016/j.eururo.2011.11.053

Brath, O. B. (2020). *Learning from incidents: Evaluating safety interventions after incident investigations* [University of Stavanger, Norway]. https://uis.brage.unit.no/uis-xmlui/handle/11250/2681460

Braut, G. S., Solberg, O., Nja, O., Solberg, T., &Njå, O. (2014). Organizational effects of experience from accidents. Learning in the aftermath of the Tretten and Asta train accidents. *Transportation Research Part A: Policy and Practice*, *69*, 354–366. https://doi.org/10.1016/j.tra.2014.08.013

Broscheit, S. (2020). Investigating Entity Knowledge in BERT with Simple Neural End-To-End Entity Linking. *CoNLL 2019 - 23rd Conference on Computational Natural Language Learning, Proceedings of the Conference*, 677–685.

https://doi.org/10.18653/v1/K19-1063

Brusa, G., Caliusco, M. L., &Chiotti, O. (2006). A process for building a domain ontology: an experience in developing a government budgetary ontology. *Proceedings of the Second Australasian Workshop on Advances in Ontologies*, *72*(c), 7–15. http://dl.acm.org/citation.cfm?id=1273661

Burdzik, R., Nowak, B., Rozmus, J., Słowiñski, P., &Pankiewicz, J. (2017). Safety in the railway industry. *Archives of Transport*, *Vol. 44*, iss. *4*(4), 15–24. https://doi.org/10.5604/01.3001.0010.6158

Bureika, G., Nozhenko, V., &Prosvirova, O. (2018). Active technical means of increasing safety in railway level crossings. *Transport Means - Proceedings of the International Conference*, *2018-Octob*, 351–355. https://www.scopus.com/inward/record.uri?eid=2-s2.0-85055561479&partnerID=40&md5=d6293750296b1f11db1e9cd9cd31f2af

Cadarso, L., Marín, ángel, &Maróti, G. (2013). Recovery of disruptions in rapid transit networks. *Transportation Research Part E: Logistics and Transportation Review*, *53*(1), 15–33. https://doi.org/10.1016/j.tre.2013.01.013

Cao, T., Mu, W., Gou, J., &Peng, L. (2020). A Study of Risk Relevance Reasoning Based on a Context Ontology of Railway Accidents. *Risk Analysis*, *40*(8), 1589–1611. https://doi.org/10.1111/risa.13506

Cao, T., Mu, W., Montarnal, A., &Barthe-Delanoë, A. M. (2019). A Method of Ontology Evolution and Concept Evaluation Based on Knowledge Discovery in the Heavy Haul Railway Risk System. *IFIP Advances in Information and Communication Technology*, *568*, 220–233. https://doi.org/10.1007/978-3-030-28464-0_20

Carayon, P., Xie, A., & Kianfar, S. (2013). Human factors and ergonomics as a patient safety practice. BMJ Quality & Safety.

Castillo, E., Grande, Z., &Calviño, A. (2016). Bayesian Networks-Based Probabilistic Safety Analysis for Railway Lines. *Computer-Aided Civil and Infrastructure Engineering*, *31*(9), 681–700. https://doi.org/10.1111/mice.12195

Cats, O., Yap, M., &vanOort, N. (2016). Exposing the role of exposure: Public transport network risk analysis. *Transportation Research Part A: Policy and Practice*, *88*, 1–14. https://doi.org/10.1016/j.tra.2016.03.015

Cedergren, A. (2013). Implementing recommendations from accident investigations: A case study of inter-organisational challenges. *Accident Analysis and Prevention*, *53*, 133–141. https://doi.org/10.1016/j.aap.2013.01.010

Cedergren, A., &Petersen, K. (2010). Learning from accident investigations - A cross-country comparison. *10th International Conference on Probabilistic Safety Assessment and Management 2010, PSAM 2010*, *2*, 1693–1703. https://www.scopus.com/inward/record.uri?eid=2-s2.0-84873598544&partnerID=40&md5=ecb5bf683237e9cf0275420e5142cac5

Cedergren, A., &Petersen, K. (2011). Prerequisites for learning from accident investigations – A cross-country comparison of national accident investigation boards. *Safety Science*, *49*(8–9), 1238–1245. https://doi.org/10.1016/J.SSCI.2011.04.005

Chang, S. E., &Nojima, N. (2001). Measuring post-disaster transportation system performance: The 1995 Kobe earthquake in comparative perspective. *Transportation Research Part A: Policy and Practice*, *35*(6), 475–494. https://doi.org/10.1016/S0965-8564(00)00003-3

Chen, X., Jia, S., &Xiang, Y. (2020). A review: Knowledge reasoning over knowledge graph. *Expert Systems with Applications*, *141*, 112948. https://doi.org/10.1016/J.ESWA.2019.112948

Chen, Z., Wang, Y., Zhao, B., Cheng, J., Zhao, X., &Duan, Z. (2020). Knowledge graph completion: A review. *IEEE Access*, *8*, 192435–192456. https://doi.org/10.1109/ACCESS.2020.3030076

Cheng, M. Y., Tsai, H. C., &Chiu, Y. H. (2009). Fuzzy case-based reasoning for coping with construction disputes. *Expert Systems with Applications*, *36*(2 PART 2), 4106–4113. https://doi.org/10.1016/j.eswa.2008.03.025

Chimalakonda, S., &Nori, K.V. (2020). An ontology based modeling framework for design of educational technologies. *Smart Learning Environments*, *7*(1), 28. https://doi.org/10.1186/s40561-020-00135-6

Cho, K., VanMerriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., &Bengio, Y. (2014). Learning phrase representations using RNN encoder-decoder for statistical machine translation. *EMNLP 2014 - 2014 Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference*, 1724–1734. https://doi.org/10.3115/v1/d14-1179

Choi, J. D., Tetreault, J., &Stent, A. (2015). It Depends: Dependency Parser Comparison Using A Web-based Evaluation Tool. *ACL-IJCNLP 2015 - 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing of the Asian Federation of Natural Language Processing, Proceedings of the Conference*, *1*, 387–396. https://doi.org/10.3115/V1/P15-1038

Chomsky, N. (1956). Three models for the description of language. *IRE Transactions on*

Information Theory, *2*(3), 113–124. https://doi.org/10.1109/TIT.1956.1056813

Chung, J., Gulcehre, C., Cho, K., &Bengio, Y. (2014). *Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling*. http://arxiv.org/abs/1412.3555

Clarke, S. (1998). Safety culture on the UK railway network. *Work and Stress*, *12*(3), 285–292. https://doi.org/10.1080/02678379808256867

Clarke, S., &Clarke, S. G. (2000). Safety culture: under-specified and overrated? *International Journal of Management Reviews*, *2*(1), 65–90. https://doi.org/10.1111/1468-2370.00031

Collobert, R., Weston, J., Com, J., Karlen, M., Kavukcuoglu, K., &Kuksa, P. (2011). Natural Language Processing (Almost) from Scratch. In *Journal of Machine Learning Research* (Vol. 12, Issue 76). http://jmlr.org/papers/v12/collobert11a.html

Corcho, O., &Gómez-Pérez, A. (2000). A roadmap to ontology specification languages. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *1937*, 80–96. https://doi.org/10.1007/3-540-39967-4_7/COVER

Corsar, D., Markovic, M., Edwards, P., &Nelson, J. D. (2015). The transport disruption ontology. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *9367*, 329–336. https://doi.org/10.1007/978-3-319-25010-6_22

Cozzani, V., Spadoni, G., Giusti, S., &Zanelli, S. (2004). The use of HazOp and Fault Tree techniques for the assessment of non-accident induced release frequencies in the transport of hazardous substances. *Probabilistic Safety Assessment and Management*, 2505–2510. https://doi.org/10.1007/978-0-85729-410-4_402

D’Este, G. M., &Taylor, M. A. P. (2003). Network Vulnerability: An Approach to Reliability Analysis at the Level of National Strategic Transport Networks. In *The Network Reliability of Transport* (pp. 23–44). Emerald Group Publishing Limited. https://doi.org/10.1108/9781786359544-002

Dai, B., &Wang, T. (2010). Risk assessment based on accident theory in urban railway transportation. *Proceedings - 2010 International Conference on Intelligent System Design and Engineering Application, ISDEA 2010*, *2*, 318–320. https://doi.org/10.1109/ISDEA.2010.374

Dang, N. C., Moreno-García, M. N., &Dela Prieta, F. (2020). Sentiment Analysis Based on Deep Learning: A Comparative Study. *Electronics*, *9*(3), 483. https://doi.org/10.3390/electronics9030483

Darroch, N., Beecroft, M., &Nelson, J. D. (2016). A conceptual framework for land use and metro infrastructure. *Infrastructure Asset Management*, *3*(4), 122–131. https://doi.org/10.1680/jinam.16.00008

Darroch, N., Beecroft, M., &Nelson, J. D. (2018). Going underground: An exploration of the interfaces between underground urban transport infrastructure and its environment. *Tunnelling and Underground Space Technology*, *81*(May), 450–462. https://doi.org/10.1016/j.tust.2018.08.027

Davies, J. (2010). Lightweight ontologies. In *Theory and Applications of Ontology: Computer Applications* (pp. 197–229). Springer Netherlands. https://doi.org/10.1007/978-90-481-8847-5_9

Dawson, D., Shaw, J., &Roland Gehrels, W. (2016). Sea-level rise impacts on transport infrastructure: The notorious case of the coastal railway line at Dawlish, England. *Journal of Transport Geography*, *51*, 97–109.

https://doi.org/10.1016/j.jtrangeo.2015.11.009

DeFelice, F., &Petrillo, A. (2011). Methodological Approach for Performing Human Reliability and Error Analysis in Railway Transportation System. *Article in International Journal of Engineering and Technology*, *3*(5), 341–353. https://www.researchgate.net/publication/281176039

deMatos, P., Dekker, A., Ennis, M., Hastings, J., Haug, K., Turner, S., &Steinbeck, C. (2010). ChEBI: a chemistry ontology and database. *Journal of Cheminformatics*, *2*(S1), 1–1. https://doi.org/10.1186/1758-2946-2-s1-p6

Debbech, S., Collart-Dutilleul, S., &Bon, P. (2020). An ontological approach to support dysfunctional analysis for railway systems design. *Journal of Universal Computer Science*, *26*(5), 549–582.

Dehghani, M. S., Flintsch, G., &McNeil, S. (2014). Impact of road conditions and disruption uncertainties on network vulnerability. *Journal of Infrastructure Systems*, *20*(3). https://doi.org/10.1061/(ASCE)IS.1943-555X.0000205

Devadoss, R., Ahmad, S. S., &Raman, D. (2012). *Platform-train interface for rail passengers-a technology review*. www.railcrc.net.au

Devlin, J., Chang, M.-W., Lee, K., &Toutanova, K. (2018). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *NAACL HLT 2019 - 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Conference*, *1*, 4171–4186. http://arxiv.org/abs/1810.04805

Diab, E., &Shalaby, A. (2020). Metro transit system resilience: Understanding the impacts of outdoor tracks and weather conditions on metro system interruptions. *International*

Journal of Sustainable Transportation, *14*(9), 657–670. https://doi.org/10.1080/15568318.2019.1600174

Dornick, C., Kumar, A., Seidenberger, S., Seidle, E., &Mukherjee, P. (2021). Analysis of Patterns and Trends in COVID-19 Research. *Big Data, Iot, and Ai for a Smarter Future*, *185*, 302–310. https://doi.org/10.1016/J.PROCS.2021.05.032

Dreyer, M., Renn, O., Cope, S., &Frewer, L. J. (2010). Including social impact assessment in food safety governance. *Food Control*, *21*(12), 1620–1628. https://doi.org/10.1016/j.foodcont.2009.05.007

Drupsteen, L., &Hasle, P. (2014). Why do organizations not learn from incidents? Bottlenecks, causes and conditions for a failure to effectively learn. *Accident Analysis and Prevention*, *72*, 351–358. https://doi.org/10.1016/j.aap.2014.07.027

Du, J., Zhao, H., Yu, Y., &Hu, Q. (2020). A Method to Extract Causality for Safety Events in Chemical Accidents from Fault Trees and Accident Reports. *Computational Intelligence and Neuroscience*, *2020*. https://doi.org/10.1155/2020/7132072

Duffield, S., &Whitty, S. J. (2015). Developing a systemic lessons learned knowledge model for organisational learning through projects. *International Journal of Project Management*, *33*(2), 311–324. https://doi.org/10.1016/J.IJPROMAN.2014.07.004

Eid, M. S., &El-adaway, I. H. (2017). Sustainable Disaster Recovery: Multiagent-Based Model for Integrating Environmental Vulnerability into Decision-Making Processes of the Associated Stakeholders. *Journal of Urban Planning and Development*, *143*(1), 04016022. https://doi.org/10.1061/(asce)up.1943-5444.0000349

Elman, J. L. (1990). Finding Structure in Time. *Cognitive Science*, *14*(2), 179–211. https://doi.org/10.1207/s15516709cog1402_1

Elms, D. (2001). Rail safety. *Reliability Engineering and System Safety*, *74*(3), 291–297. https://doi.org/10.1016/S0951-8320(01)00085-0

Elsahar, H., Vougiouklis, P., Remaci, A., Gravier, C., Hare, J., Laforest, F., &Simperl, E. (2018). T-REx: A Large Scale Alignment of Natural Language with Knowledge Base Triples. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation*. https://aclanthology.org/L18-1544

Ettinger, S., Mounaud, L., Magill, C., Yao-Lafourcade, A. F., Thouret, J. C., Manville, V., Negulescu, C., Zuccaro, G., DeGregorio, D., Nardone, S., Uchuchoque, J. A. L., Arguedas, A., Macedo, L., &Manrique Llerena, N. (2016). Building vulnerability to hydro-geomorphic hazards: Estimating damage probability from qualitative vulnerability assessment using logistic regression. *Journal of Hydrology*, *541*, 563–581. https://doi.org/10.1016/j.jhydrol.2015.04.017

European Railway Agency (ERA). (2007). Safety Management System– Assessment Criteria for Railway Undertakings and Infrastructure Managers (ERA Working Group on Safety Certification and Authorisation ed.).

Evans, A. W. (2011). Fatal train accidents on Europe’s railways: 1980-2009. *Accident Analysis and Prevention*, *43*(1), 391–401. https://doi.org/10.1016/j.aap.2010.09.009

Fan, Y., Li, Z., Pei, J., Li, H., &Sun, J. (2015). Applying systems thinking approach to accident analysis in China: Case study of “7.23” Yong-Tai-Wen High-Speed train accident. *Safety Science*, *76*, 190–201. https://doi.org/10.1016/j.ssci.2015.02.017

Färber, M., Bartscherer, F., Menne, C., &Rettinger, A. (2018). Linked data quality of DBpedia, Freebase, OpenCyc, Wikidata, and YAGO. *Semantic Web*, *9*(1), 77–129. https://doi.org/10.3233/SW-170275

Ferragina, P., &Scaiella, U. (2010). TAGME: On-the-fly annotation of short text fragments (by Wikipedia entities). *International Conference on Information and Knowledge Management, Proceedings*, 1625–1628. https://doi.org/10.1145/1871437.1871689

Figueres-Esteban, M., Hughes, P., &vanGulijk, C. (2016). Visual analytics for text-based railway incident reports. *SAFETY SCIENCE*, *89*, 72–76. https://doi.org/10.1016/j.ssci.2016.05.009

Filho, A. P. G., Ferreira, A. M. S., Ramos, M. F., &Pinto, A. R. A. P. (2021). Are we learning from disasters? Examining investigation reports from National government bodies. *Safety Science*, *140*, 105327. https://doi.org/10.1016/J.SSCI.2021.105327

Fischbach, J., Hauptmann, B., Konwitschny, L., Spies, D., &Vogelsang, A. (2020). Towards Causality Extraction from Requirements. *Proceedings of the IEEE International Conference on Requirements Engineering*, *2020-August*, 388–393. https://doi.org/10.1109/RE48521.2020.00053

FRA. (2010). *“About the FRA.”*

FRA. (2020). *HQ-2020-1377*. https://railroads.dot.gov/elibrary/hq-2020-1377

Frey, J., Hofer, M., Obraczka, D., Lehmann, J., &Hellmann, S. (2019). DBpedia FlexiFusion the Best of Wikipedia > Wikidata > Your Data. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *11779 LNCS*, 96–112. https://doi.org/10.1007/978-3-030-30796-7_7

Ganesh, P., Chen, Y., Lou, X., Khan, M. A., Yang, Y., Chen, D., Winslett, M., Sajjad, H., &Nakov, P. (2020). Compressing Large-Scale Transformer-Based Models: A Case Study on BERT. *ArXiv*. http://arxiv.org/abs/2002.11985

Georges, L. R. A., &vanWitteloostuijn, A. (1999). Circular organizing and triple loop learning. *Journal of Organizational Change Management*, *12*(5), 439–454. https://doi.org/10.1108/09534819910289110

Gers, F. A. (1999). Learning to forget: continual prediction with LSTM. *9th International Conference on Artificial Neural Networks: ICANN '99*, *1999*(470), 850–855. https://doi.org/10.1049/cp:19991218

Goldbeck, N., Angeloudis, P., &Ochieng, W. (2020). Optimal supply chain resilience with consideration of failure propagation and repair logistics. *Transportation Research Part E: Logistics and Transportation Review*, *133*, 101830. https://doi.org/10.1016/j.tre.2019.101830

Gong, Y., &Li, Y. (2018). STAMP-based causal analysis of China-Donghuang oil transportation pipeline leakage and explosion accident. *Journal of Loss Prevention in the Process Industries*, *56*, 402–413. https://doi.org/10.1016/j.jlp.2018.10.001

Goodfellow, I., Bengio, Y., &Courville, A. (2016). Deep learning. In *Deep learning*. MIT press.

Grant, E., Salmon, P. M., Stevens, N. J., Goode, N., &Read, G. J. (2016). Exposing resistant problems in complex systems: A review of accident causation tenets. *51st Annual Conference of the Human Factors and Ergonomics Society of Australia 2016: Healthy, Safe and Productive By Design*, 35–49. https://www.scopus.com/inward/record.uri?eid=2-s2.0-85018425920&partnerID=40&md5=41fa56ac2a9b20ccb422c0cb0fe44ba7

Graves, A. (2013). *Generating Sequences With Recurrent Neural Networks*. http://arxiv.org/abs/1308.0850

Greff, K., Srivastava, R. K., Koutnik, J., Steunebrink, B. R., &Schmidhuber, J. (2017).

LSTM: A Search Space Odyssey. *IEEE Transactions on Neural Networks and Learning Systems*, *28*(10), 2222–2232. https://doi.org/10.1109/TNNLS.2016.2582924

Grootendorst, M. (2022). *BERTopic: Neural topic modeling with a class-based TF-IDF procedure*. https://doi.org/10.48550/arxiv.2203.05794

Grüninger, M., Grüninger, M., &Fox, M. S. (1995). *Methodology for the Design and Evaluation of Ontologies*. http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.44.8723

Gudivada, V. N., &Arbabifard, K. (2018). Open-Source Libraries, Application Frameworks, and Workflow Systems for NLP. In *Handbook of Statistics* (Vol. 38, pp. 31–50). Elsevier B.V. https://doi.org/10.1016/bs.host.2018.07.007

Guo, C., Lu, M., &Wei, W. (2019). An Improved LDA Topic Modeling Method Based on Partition for Medium and Long Texts. *Annals of Data Science 2019 8:2*, *8*(2), 331–344. https://doi.org/10.1007/S40745-019-00218-3

Gurulingappa, H., Rajput, A. M., Roberts, A., Fluck, J., Hofmann-Apitius, M., &Toldo, L. (2012). Development of a benchmark corpus to support the automatic extraction of drug-related adverse effects from medical case reports. *Journal of Biomedical Informatics*, *45*(5), 885–892. https://doi.org/10.1016/j.jbi.2012.04.008

Hadj-Mabrouk, H. (2020). Analysis and prediction of railway accident risks using machine learning. *AIMS Electronics and Electrical Engineering*, *4*(1), 19–46. https://doi.org/10.3934/electreng.2020.1.19

Haendel, M. A., Chute, C. G., &Robinson, P. N. (2018). Classification, Ontology, and Precision Medicine. *New England Journal of Medicine*, *379*(15), 1452–1462. https://doi.org/10.1056/nejmra1615014

Haimes, Y. Y. (2006). On the definition of vulnerabilities in measuring risks to infrastructures. In *Risk Analysis* (Vol. 26, Issue 2, pp. 293–296). https://doi.org/10.1111/j.1539-6924.2006.00755.x

Hakim, A. A., Erwin, A., Eng, K. I., Galinium, M., &Muliady, W. (2014, January 12). Automated document classification for news article in Bahasa Indonesia based on term frequency inverse document frequency (TF-IDF) approach. *Proceedings - 2014 6th International Conference on Information Technology and Electrical Engineering: Leveraging Research and Technology Through University-Industry Collaboration, ICITEE 2014*. https://doi.org/10.1109/ICITEED.2014.7007894

Han, X., &Eisenstein, J. (2019). Unsupervised Domain Adaptation of Contextualized Embeddings for Sequence Labeling. *EMNLP-IJCNLP 2019 - 2019 Conference on Empirical Methods in Natural Language Processing and 9th International Joint Conference on Natural Language Processing, Proceedings of the Conference*, 4238–4248. https://doi.org/10.18653/V1/D19-1433

Hansen, D. L., Shneiderman, B., Smith, M. A., &Himelboim, I. (2019). Calculating and visualizing network metrics. In *Analyzing Social Media Networks with NodeXL: Insights from a Connected World*. Elsevier. https://doi.org/10.1016/B978-0-12-817756-3.00006-6

Harris, Z. S. (1954). Distributional Structure. WORD, *10*(2–3), 146–162. https://doi.org/10.1080/00437956.1954.11659520

Hasibi, F., Balog, K., &Bratsberg, S. E. (2016). On the Reproducibility of the TAGME Entity Linking System. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *9626*, 436–449. https://doi.org/10.1007/978-3-319-30671-1_32

Hedlund, F. H., Nielsen, M. F., Mikkelsen, S. H., &Kragh, E. K. (2014). Violent explosion after inadvertent mixing of nitric acid and isopropanol – Review 15 years later finds basic accident data corrupted, no evidence of broad learning. *Safety Science*, *70*, 255–261. https://doi.org/10.1016/J.SSCI.2014.06.010

Heidarysafa, M., Kowsari, K., Barnes, L. E., &Brown, D. E. (2019). Analysis of Railway Accidents' Narratives Using Deep Learning. *Proceedings - 17th IEEE International Conference on Machine Learning and Applications, ICMLA 2018*, *17th IEEE International Conference on Machine Learning and Applications (IEEE ICMLA)*, 1446–1453. https://doi.org/10.1109/ICMLA.2018.00235

Heindorf, S., Scholten, Y., Wachsmuth, H., Ngonga Ngomo, A. C., &Potthast, M. (2020). CauseNet: Towards a Causality Graph Extracted from the Web. *International Conference on Information and Knowledge Management, Proceedings*, 3023–3030. https://doi.org/10.1145/3340531.3412763

Heinrich, H. W. (1941). Industrial Accident Prevention. A Scientific Approach. In *cabdirect.org* (Second Edition). Industrial Accident Prevention. A Scientific Approach. https://www.cabdirect.org/cabdirect/abstract/19432701767

Hendrickx, I., Kim, S. N., Kozareva, Z., Nakov, P., Diarmuid´o, D., Diarmuid´o, S. ¶, Padó, S., Pennacchiotti, M., Romano, L., &Szpakowicz, S. (2010). SemEval-2010 Task 8: Multi-Way Classification of Semantic Relations Between Pairs of Nominals. *ACL 2010 - SemEval 2010 - 5th International Workshop on Semantic Evaluation, Proceedings*, 33–38.

Hochreiter, S., &Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, *9*(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

Holmes, M., &Robertson, N. (2021). Promoting Workforce Planning as a Means of

Embedding a Learning Culture in the Rail Sector. *Education in the Asia-Pacific Region*, *58*, 227–240. https://doi.org/10.1007/978-981-16-0983-1_16

Holmgren, Å. (2004). *Vulnerability Analysis of Electric Power Delivery Networks*. https://www.diva-portal.org/smash/record.jsf?pid=diva2:7703

Hong, L., Ouyang, M., Peeta, S., He, X., &Yan, Y. (2015). Vulnerability assessment and mitigation for the Chinese railway system under floods. *Reliability Engineering and System Safety*, *137*, 58–68. https://doi.org/10.1016/j.ress.2014.12.013

Hong, L., Yan, Y., Ouyang, M., Tian, H., &He, X. (2017). Vulnerability effects of passengers' intermodal transfer distance preference and subway expansion on complementary urban public transportation systems. *Reliability Engineering and System Safety*, *158*, 58–72. https://doi.org/10.1016/j.ress.2016.10.001

Hong, W.-T., Clifton, G., &Nelson, J. D. (2022). Rail transport system vulnerability analysis and policy implementation: Past progress and future directions. *Transport Policy*, *128*, 299–308. https://doi.org/https://doi.org/10.1016/j.tranpol.2022.02.004

Hong, W.-T., Clifton, G., &Nelson, J. D. (2023). Railway accident causation analysis: Current approaches, challenges and potential solutions. *Accident Analysis & Prevention*, *186*, 107049. https://doi.org/https://doi.org/10.1016/j.aap.2023.107049

Howard, J., &Gugger, S. (2020). Fastai: A Layered API for Deep Learning. *Information*, *11*(2), 108. https://doi.org/10.3390/info11020108

Hua, L., Zheng, W., &Gao, S. (2019). Extraction and Analysis of Risk Factors from Chinese Railway Accident Reports. *2019 IEEE Intelligent Transportation Systems Conference, ITSC 2019*, 869–874. https://doi.org/10.1109/ITSC.2019.8917094

Hua, W., &Ong, G. P. (2017). Network survivability and recoverability in urban rail transit systems under disruption. *IET Intelligent Transport Systems*, *11*(10), 641–648. https://doi.org/10.1049/iet-its.2017.0102

Huang, W., Liu, Y., Zhang, Y., Zhang, R., Xu, M., DeDieu, G. J., Antwi, E., &Shuai, B. (2020). Fault Tree and Fuzzy D-S Evidential Reasoning combined approach: An application in railway dangerous goods transportation system accident analysis. *Information Sciences*, *520*, 117–129. https://doi.org/10.1016/j.ins.2019.12.089

Huang, W., Shuai, B., Zhang, R., Xu, M., Xu, Y., Yu, Y., &Antwi, E. (2020). A New System Risk Definition and System Risk Analysis Approach Based on Improved Risk Field. *IEEE Transactions on Reliability*, *69*(4), 1437–1452. https://doi.org/10.1109/TR.2019.2942373

Huang, W., Shuai, B., Zuo, B., Xu, Y., &Antwi, E. (2019). A systematic railway dangerous goods transportation system risk analysis approach: The 24 model. *Journal of Loss Prevention in the Process Industries*, *61*, 94–103. https://doi.org/10.1016/J.JLP.2019.05.021

Huggins, R., &Johnston, A. (2010). Knowledge flow and inter-firm networks: The influence of network resources, spatial proximity and firm size. *Http://Dx.Doi.Org/10.1080/08985620903171350*, *22*(5), 457–484. https://doi.org/10.1080/08985620903171350

Hughes, P., Robinson, R., Figueres-Esteban, M., &vanGulijk, C. (2019). Extracting safety information from multi-lingual accident reports using an ontology-based approach. *Safety Science*, *118*(May), 288–297. https://doi.org/10.1016/j.ssci.2019.05.029

Hulin, B., Kaindl, H., Rathfux, T., Popp, R., Arnautovic, E., &Beckert, R. (2016). Towards a Common Safety Ontology for Automobiles and Railway Vehicles. *Proceedings - 2016 12th European Dependable Computing Conference, EDCC 2016*, 189–192.

https://doi.org/10.1109/EDCC.2016.15

Hulme, A., Stanton, N. A., Walker, G. H., Waterson, P., &Salmon, P. M. (2019). What do applications of systems thinking accident analysis methods tell us about accident causation? A systematic review of applications between 1990 and 2018. *Safety Science*, *117*, 164–183. https://doi.org/10.1016/J.SSCI.2019.04.016

Hutchings, J., &Thatcher, A. (2019). A systems analysis of the South African railway industry. *Advances in Intelligent Systems and Computing*, *823*, 305–313. https://doi.org/10.1007/978-3-319-96074-6_33/FIGURES/4

International Civil Aviation Organization (ICAO). (1944). *About ICAO*. https://www.icao.int/about-icao/Pages/default.aspx

International Civil Aviation Organization (ICAO). (2007). Dangerous Goods Panel (DGP) Meeting of the Working Group of the Whole (DGP-WG/07-WP/58). Retrieved from the ICAO:<https://www.icao.int/safety/DangerousGoods/Working%20Group %20of%20the%20Whole%2007/DGPWG.07.WP.058.5.en.pdf >.

Immers, B., Yperman, I., Stada, J., &Bleukx, A. (2004). Reliability and robustness of transportation networks; problem survey and examples. *Proceedings of the NECTAR Cluster Meeting on Reliability of Networks*. https://lirias.kuleuven.be/1749567

Issa, S., Adekunle, O., Hamdi, F., Cherfi, S. S. S., Dumontier, M., &Zaveri, A. (2021). Knowledge Graph Completeness: A Systematic Literature Review. *IEEE Access*, *9*, 31322–31339. https://doi.org/10.1109/ACCESS.2021.3056622

Jeffcott, S., Pidgeon, N., Weyman, A., &Walls, J. (2006). Risk, trust, and safety culture in U.K. train operating companies. *Risk Analysis*, *26*(5), 1105–1121.

https://doi.org/10.1111/j.1539-6924.2006.00819.x

Jenelius, E., Petersen, T., &Mattsson, L. G. (2006). Importance and exposure in road network vulnerability analysis. *Transportation Research Part A: Policy and Practice*, *40*(7), 537–560. https://doi.org/10.1016/j.tra.2005.11.003

Jiang, Y., &Kong, D. (2013). Research of collaborative management of organization interface of large-scale engineering project. *Proceedings - 2013 International Conference on Information Technology and Applications, ITA 2013*, 450–453. https://doi.org/10.1109/ITA.2013.109

Jin, J. G., Tang, L. C., Sun, L., &Lee, D. H. (2014). Enhancing metro network resilience via localized integration with bus services. *Transportation Research Part E: Logistics and Transportation Review*, *63*, 17–30. https://doi.org/10.1016/j.tre.2014.01.002

Johnsen, S. O., Vatn, J., Rosness, R., &Herrera, I. A. (2006). Cross border railway operations: Improving safety at cultural interfaces. *Cognition, Technology and Work*, *8*(1), 76–88. https://doi.org/10.1007/s10111-005-0022-8

Johnson, W. (1973). *The Management Oversight and Risk Tree-MORT* (Vol. 7). https://books.google.com/books?hl=zh-TW&lr=&id=EolHAQAAMAAJ&oi=fnd&pg=PA1&dq=Management+oversight+and+risk+tree+&ots=-ipfTqMHdq&sig=rx9ip3v-iSnJNzjwoC4oiSNWu4M

Jones, K. S. (1972). A statistical interpretation of term specificity and its application in retrieval. In *Journal of Documentation* (Vol. 28, Issue 1, pp. 11–21). MCB UP Ltd. https://doi.org/10.1108/eb026526

Jones, R. N. (2001). An environmental risk assessment/management framework for climate change impact assessments. *Natural Hazards*, *23*(2–3), 197–230.

https://doi.org/10.1023/A:1011148019213

Jugran, S., Kumar, A., Tyagi, B. S., &Anand, V. (2021). Extractive Automatic Text Summarization using SpaCy in Python NLP. *2021 International Conference on Advance Computing and Innovative Technologies in Engineering, ICACITE 2021*, 582–585. https://doi.org/10.1109/ICACITE51222.2021.9404712

Kalem, A., Lindov, O., &Šimić, E. (2021). Safety Culture in the Function of Optimization of Railway Safety Management System. In *Lecture Notes in Networks and Systems* (Vol. 233). https://doi.org/10.1007/978-3-030-75275-0_77

Kaplan, S., &Garrick, B. J. (1981). On The Quantitative Definition of Risk. *Risk Analysis*, *1*(1), 11–27. https://doi.org/10.1111/j.1539-6924.1981.tb01350.x

Karanikas, N. (2016). Exploiting data from safety investigations and processes to assess performance of safety management aspects. *Http://Dx.Doi.Org/10.1080/14773996.2016.1255444*, *14*(2), 115–127. https://doi.org/10.1080/14773996.2016.1255444

Karimi, I., &Hüllermeier, E. (2007). Risk assessment system of natural hazards: A new approach based on fuzzy probability. *Fuzzy Sets and Systems*, *158*(9), 987–999. https://doi.org/10.1016/j.fss.2006.12.013

Katsumi, M., &Fox, M. (2018). Ontologies for Transportation Research: A Survey. *Transportation Research Part C: Emerging Technologies*, *89*(September 2017), 53–82. https://doi.org/10.1016/j.trc.2018.01.023

Kaviani, A., Thompson, R. G., &Rajabifard, A. (2017). Improving regional road network resilience by optimised traffic guidance. *Transportmetrica A: Transport Science*, *13*(9), 794–828. https://doi.org/10.1080/23249935.2017.1335807

Kejriwal, M. (2019). *Domain-Specific Knowledge Graph Construction*. https://doi.org/10.1007/978-3-030-12375-8

Kelly, B., &Berger, S. (2006). Interface management: Effective communication to improve process safety. *Journal of Hazardous Materials*, *130*(3 SPEC. ISS.), 321–325. https://doi.org/10.1016/j.jhazmat.2005.07.009

Khaled, A. A., Jin, M., Clarke, D. B., &Hoque, M. A. (2015). Train design and routing optimization for evaluating criticality of freight railroad infrastructures. *Transportation Research Part B: Methodological*, *71*, 71–84. https://doi.org/10.1016/j.trb.2014.10.002

Khan, F., Abunada, H., John, D., &Benmosbah, T. (2010). Development of risk-based process safety indicators. *Process Safety Progress*, *29*(2), 133–143. https://doi.org/10.1002/prs.10354

Khetan, V., Ramnani, R., Anand, M., Sengupta, S., &Fano, A. E. (2020). *Causal BERT : Language models for causality detection between events expressed in text*. http://arxiv.org/abs/2012.05453

Khoudour, L., El-Koursi, E. M., Velastin, S. A., Buch, N., Lim-Thiebot, S., &Fontaine, F. (2011). An approach for protecting transport infrastructure. *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, *225*(4), 383–393. https://doi.org/10.1177/2041301710395076

Kim, D. S., &Yoon, W. C. (2013). An accident causation model for the railway industry: Application of the model to 80 rail accident investigation reports from the UK. *Safety Science*, *60*, 57–68. https://doi.org/10.1016/j.ssci.2013.06.010

Kim, H., &Song, Y. (2018). An integrated measure of accessibility and reliability of mass transit systems. *Transportation*, *45*(4), 1075–1100. https://doi.org/10.1007/s11116-018-

9866-7

Kim, S., &Yeo, H. (2017). Evaluating link criticality of road network based on the concept of macroscopic fundamental diagram. *Transportmetrica A: Transport Science*, *13*(2), 162–193. https://doi.org/10.1080/23249935.2016.1231231

Kinnersley, S., &Roelen, A. (2007). The contribution of design to accidents. *Safety Science*, *45*(1–2), 31–60. https://doi.org/10.1016/j.ssci.2006.08.010

Kojima, S., Mimura, Y., &Kazama, H. (2009). Quality assurance activity focused on achieving nuclear safety. *Hitachi Review*, *58*(2), 119–124.

Konys, A. (2018). Knowledge Systematization for ontology learning methods. *Procedia Computer Science*, *126*, 2194–2207. https://doi.org/10.1016/j.procS.2018.07.229

Kotis, K., &Vouros, G. A. (2005). Human-centered ontology engineering: The HCOME methodology. *Knowledge and Information Systems 2005 10:1*, *10*(1), 109–131. https://doi.org/10.1007/S10115-005-0227-4

Krdzavac, N., Mosbach, S., Nurkowski, D., Buerger, P., Akroyd, J., Martin, J., Menon, A., &Kraft, M. (2019). An Ontology and Semantic Web Service for Quantum Chemistry Calculations. *Journal of Chemical Information and Modeling*, *59*(7), 3154–3165. https://doi.org/10.1021/acs.jcim.9b00227

Krishnan, K., &Rogers, S. P. (2015). A New Universe of Data. In *Social Data Analytics* (pp. 1–10). Elsevier. https://doi.org/10.1016/b978-0-12-397186-9.00001-7

Kuhn, K. D. (2018). Using structural topic modeling to identify latent topics and trends in aviation incident reports. *Transportation Research Part C: Emerging Technologies*, *87*, 105–122. https://doi.org/10.1016/J.TRC.2017.12.018

Kume, S., &Kozaki, K. (2021). *Extracting Domain-specific Concepts from Large-scale*

Linked Open Data; Extracting Domain-specific Concepts from Large-scale Linked Open Data. https://doi.org/10.1145/3502223.3502227

Kwayu, K. M., Kwigizile, V., Lee, K., &Oh, J. S. (2021). Discovering latent themes in traffic fatal crash narratives using text mining analytics and network topology. *Accident Analysis & Prevention*, *150*, 105899. https://doi.org/10.1016/J.AAP.2020.105899

Kyriakidis, M., Pak, K. T., &Majumdar, A. (2015). Railway Accidents Caused by Human Error: Historic Analysis of UK Railways, 1945 to 2012. *Transportation Research Record: Journal of the Transportation Research Board*, *2476*(1), 126–136. https://doi.org/10.3141/2476-17

Labusch, K., &Neudecker, C. (2020). Named Entity Disambiguation and Linking on Historic Newspaper OCR with BERT. *Working Notes of CLEF 2020 - Conference and Labs of the Evaluation Forum*. https://qurator.ai

Lamb, R., Garside, P., Pant, R., &Hall, J. W. (2019). A Probabilistic Model of the Economic Risk to Britain's Railway Network from Bridge Scour During Floods. *Risk Analysis*, *39*(11), 2457–2478. https://doi.org/10.1111/risa.13370

Lämmel, G., Rieser, M., Nagel, K., Taubenböck, H., Strunz, G., Goseberg, N., Schlurmann, T., Klüpfel, H., Setiadi, N., &Birkmann, J. (2010). Emergency Preparedness in the Case of a Tsunami—Evacuation Analysis and Traffic Optimization for the Indonesian City of Padang. In *Pedestrian and Evacuation Dynamics 2008* (pp. 171–182). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-04504-2_13

Lasisi, A., &Attoh-Okine, N. (2020). An Unsupervised Learning Framework for Track Quality Index and Safety. *Transportation Infrastructure Geotechnology*, *7*(1), 1–12. https://doi.org/10.1007/S40515-019-00087-6/TABLES/2

Lata, K., Singh, P., &Dutta, K. (2022). Mention detection in coreference resolution: survey. *Applied Intelligence*, *52*(9), 9816–9860. https://doi.org/10.1007/S10489-021-02878-2/TABLES/17

Lauby, R. (2016). *Federal Register :: Mitigation and Investigation of Passenger Rail Human Factor Related Accidents and Operations in Terminals and Stations With Stub End Tracks*. https://www.federalregister.gov/documents/2016/12/05/2016-29013/mitigation-and-investigation-of-passenger-rail-human-factor-related-accidents-and-operations-in

Laurentius, G. (1994). The vulnerability of the city. *Planning a High Resilience Society. Swedish Agency for Civil Emergency and Planning (ÖCB) and Umeå Universitet. Geographical Reports*, *11*.

Lee, J. S., Min Kim, H., IlKim, S., &Min Lee, H. (2021). Evaluation of structural integrity of railway bridge using acceleration data and semi-supervised learning approach. *Engineering Structures*, *239*, 112330. https://doi.org/10.1016/J.ENGSTRUCT.2021.112330

Lehmann, J., &Völker, J. (2014). *Perspectives on ontology learning (Vol. 18).* IOS Press.

Lemos, F. (2020). On the definition of risk. *Systems Engineering and Electronics*, *13*(3), 266–278. https://doi.org/10.3969/j.issn.1001-506X.2019.03.24

Leveson, N. (2004). A new accident model for engineering safer systems. *Safety Science*, *42*(4), 237–270. https://doi.org/10.1016/S0925-7535(03)00047-X

Lewis, R. (2015). *A semantic approach to railway data integration and decision support*. PhD Thesis . http://etheses.bham.ac.uk/id/eprint/5959

Li, S.Bin, Wu, J. J., Gao, Z. Y., Lin, Y., &Fu, B. B. (2011). The analysis of traffic congestion and dynamic propagation properties based on complex network. *Wuli Xuebao/Acta*

Physica Sinica, *60*(5), 050701–050701. https://doi.org/10.7498/aps.60.050701

Li, C., Lu, Y., Wu, J., Zhang, Y., Xia, Z., Wang, T., Yu, D., Chen, X., Liu, P., Guo, J., Li, C., Lu, Y., Wu, J., Zhang, Y., Xia, Z., Wang, T., Yu, D., Chen, X., &Liu, P. (2018). LDA Meets Word2Vec: A Novel Model for Academic Abstract Clustering. *WWW '18: Companion Proceedings of the The Web Conference 2018*, 1699–1706. https://doi.org/10.1145/3184558.3191629

Li, H., Li, S., Sun, J., Xing, Z., Peng, X., Liu, M., &Zhao, X. (2018). Improving api caveats accessibility by mining api caveats knowledge graph. *Proceedings - 2018 IEEE International Conference on Software Maintenance and Evolution, ICSME 2018*, 183–193. https://doi.org/10.1109/ICSME.2018.00028

Li, J., &Guo, H. (2011). *Management & Engineering Research on Comprehensive Evaluation Index System of Organization Interface Management of Construction Project*. *05*, 99–104.

Li, J., Sun, A., Han, J., &Li, C. (2020). A Survey on Deep Learning for Named Entity Recognition. *IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING*. http://neuroner.com/

Li, X. Q., Shi, T. Y., Li, P., Yang, L. B., Ma, X. N., &ACM. (2018). BiLSTM-CRF model for named entity recognition in railway accident and fault analysis report. *ACM International Conference Proceeding Series*, *Part F1482*(Asia-Pacific Conference on Intelligent Medical (APCIM) / 7th International Conference on Transportation and Traffic Engineering (ICTTE)), 1–5. https://doi.org/10.1145/3321619.3321623

Li, Y., & Guldenmund, F. W. (2018). Safety management systems: A broad overview of the literature. Safety Science, 103, 94-123.

Liang, C., Ghazel, M., &Cazier, O. (2018). *Using Bayesian Networks for the Purpose of Risk Analysis at Railway Level Crossings*. *51*(9), 142–149. https://doi.org/10.1016/j.ifacol.2018.07.024

Liao, T. Y., Hu, T. Y., &Ko, Y. N. (2018). A resilience optimization model for transportation networks under disasters. *Natural Hazards*, *93*(1), 469–489. https://doi.org/10.1007/s11069-018-3310-3

Lin, C., Zhang, M., Zhou, Z., Li, L., Shi, S., Chen, Y., &Dai, W. (2020). A new quantitative method for risk assessment of water inrush in karst tunnels based on variable weight function and improved cloud model. *Tunnelling and Underground Space Technology*, *95*. https://doi.org/10.1016/j.tust.2019.103136

Lin, F., Wu, P., &Xu, Y. (2021). Investigation of Factors Influencing the Construction Safety of High-Speed Railway Stations Based on DEMATEL and ISM. *Advances in Civil Engineering*, *2021*. https://doi.org/10.1155/2021/9954018

Lindgren, J., Jonsson, D. K., &Carlsson-Kanyama, A. (2009). Climate adaptation of railways: Lessons from Sweden. *European Journal of Transport and Infrastructure Research*, *9*(2), 164–181. https://doi.org/10.18757/ejtir.2009.9.2.3295

Ling, L., David Gomersall, C., Samy, W., Joynt, G. M., Leung, C. C. H., Wong, W. T., &Lee, A. (2016). The Effect of a Freely Available Flipped Classroom Course on Health Care Worker Patient Safety Culture: A Prospective Controlled Study. *J Med Internet Res 2016;18(7):E180 Https://Www.Jmir.Org/2016/7/E180*, *18*(7), e5378. https://doi.org/10.2196/JMIR.5378

Lingegård, S., &Lindahl, M. (2015). Integrated Product Service Offerings for rail infrastructure – benefits and challenges regarding knowledge transfer and cultural change in a Swedish case. *Journal of Cleaner Production*, *98*, 166–174.

https://doi.org/10.1016/J.JCLEPRO.2014.06.039

Little, R. G. (2002). Controlling cascading failure: Understanding the vulnerabilities of interconnected infrastructures. *Journal of Urban Technology*, *9*(1), 109–123. https://doi.org/10.1080/106307302317379855

Liu, J., Schmid, F., Li, K., &Zheng, W. (2021a). A knowledge graph-based approach for exploring railway operational accidents. *Reliability Engineering and System Safety*, *207*. https://doi.org/10.1016/j.ress.2020.107352

Liu, J., Schmid, F., Li, K., &Zheng, W. (2021b). A knowledge graph-based approach for exploring railway operational accidents. *Reliability Engineering & System Safety*, *207*, 107352. https://doi.org/10.1016/J.RESS.2020.107352

Liu, P., Yang, L., Gao, Z., Li, S., &Gao, Y. (2015). Fault tree analysis combined with quantitative analysis for high-speed railway accidents. *Safety Science*, *79*, 344–357. https://doi.org/10.1016/j.ssci.2015.06.017

Lovrencic, S., &Cubrilo, M. (2008). Ontology Evaluation - Comprising Verification and Validation - ProQuest. *Central European Conference on Information and Intelligent Systems*, 1–7. https://www.proquest.com/docview/1314945407?pq-origsite=gscholar&fromopenview=true

Luhn, H. P. (1957). A Statistical Approach to Mechanized Encoding and Searching of Literary Information. *IBM Journal of Research and Development*, *1*(4), 309–317. https://doi.org/10.1147/rd.14.0309

Lund, K., &Burgess, C. (1996). Producing high-dimensional semantic spaces from lexical co-occurrence. *Behavior Research Methods, Instruments, and Computers*, *28*(2), 203–208. https://doi.org/10.3758/BF03204766

Luo, L., Yang, Z., Yang, P., Zhang, Y., Wang, L., Lin, H., &Wang, J. (2018). An attention-based BiLSTM-CRF approach to document-level chemical named entity recognition. *Bioinformatics*, *34*(8), 1381–1388. https://doi.org/10.1093/bioinformatics/btx761

Ly, A., Uthayasooriyar, B., &Wang, T. (2020). A survey on natural language processing (nlp) and applications in insurance. *ArXiv*. http://arxiv.org/abs/2010.00462

Maalel, A., Mejri, L., Mabrouk, H. H., &BenGhezela, H. (2012a). Toward a knowledge management approach based on an ontology and Case-based Reasoning (CBR): Application to railroad accidents. *Proceedings - International Conference on Research Challenges in Information Science*, *May 2016*. https://doi.org/10.1109/RCIS.2012.6240448

Maalel, A., Mejri, L., Mabrouk, H. H., &BenGhezela, H. (2012b). Towards an ontology of help to the modeling of accident scenarii: Application on railroad transport. *2012 6th International Conference on Sciences of Electronics, Technologies of Information and Telecommunications, SETIT 2012*, *Figure 1*, 1–6. https://doi.org/10.1109/SETIT.2012.6481880

Madigan, R., Golightly, D., &Madders, R. (2016). Application of Human Factors Analysis and Classification System (HFACS) to UK rail safety of the line incidents. *Accident Analysis and Prevention*, *97*, 122–131. https://doi.org/10.1016/j.aap.2016.08.023

Magnini, B., Lavelli, A., Fondazione, S. M., &Kessler, B. (2020). *Comparing Machine Learning and Deep Learning Approaches on NLP Tasks for the Italian Language*. https://github.com/google-research/bert/

Manning, C. D., Surdeanu, M., Bauer, J., Finkel, J. R., Bethard, S., &McClosky, D. (2014). *The Stanford CoreNLP Natural Language Processing Toolkit*. 55–60. https://doi.org/10.3115/V1/P14-5010

Marzuoli, A., Boidot, E., Colomar, P., Guerpillon, M., Feron, E., Bayen, A., &Hansen, M. (2016). Improving Disruption Management with Multimodal Collaborative Decision-Making: A Case Study of the Asiana Crash and Lessons Learned. *IEEE Transactions on Intelligent Transportation Systems*, *17*(10), 2699–2717. https://doi.org/10.1109/TITS.2016.2536733

Mathews, S. M. (2019). Explainable Artificial Intelligence Applications in NLP, Biomedical, and Malware Classification: A Literature Review. *Advances in Intelligent Systems and Computing*, *998*, 1269–1292. https://doi.org/10.1007/978-3-030-22868-2_90

Matthew English, S. (2018). An Extensible Schema for Building Large Weakly-Labeled Semantic Corpora. *Procedia Computer Science*, *128*, 65–71. https://doi.org/10.1016/J.PROCS.2018.03.009

Mattsson, L. G., &Jenelius, E. (2015). Vulnerability and resilience of transport systems - A discussion of recent research. *Transportation Research Part A: Policy and Practice*, *81*, 16–34. https://doi.org/10.1016/j.tra.2015.06.002

McDaniels, T., Chang, S., Cole, D., Mikawoz, J., &Longstaff, H. (2008). Fostering resilience to extreme events within infrastructure systems: Characterizing decision contexts for mitigation and adaptation. *Global Environmental Change*, *18*(2), 310–318. https://doi.org/10.1016/j.gloenvcha.2008.03.001

McGuinness, D. L., &Harmelen, F.van. (2004). OWL Web Ontology Language - Reference. *W3C Recommendation [Online], Available at: Http://Www.W3.Org/TR/2004/REC-Owl*, 1–53, [Accessed: 02 February 2016]. https://research.vu.nl/en/publications/owl-web-ontology-language-reference-2

McHugh, K., &Klockner, K. (2020). Learning lessons from rail safety storytelling: Telling safety like it is. *Safety Science*, *122*, 104524.

https://doi.org/10.1016/J.SSCI.2019.104524

McInnes, L., Healy, J., &Astels, S. (2017). hdbscan: Hierarchical density based clustering. *Journal of Open Source Software*, *2*(11), 205. https://doi.org/10.21105/JOSS.00205

McInnes, L., Healy, J., &Melville, J. (2018). *UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction*. https://doi.org/10.48550/arxiv.1802.03426

Meng, Y., Zhang, Y., Huang, J., Xiong, C., Ji, H., Zhang, C., &Han, J. (2020). Text classification using label names only: A language model self-training approach. In *arXiv*. arXiv. https://doi.org/10.18653/v1/2020.emnlp-main.724

Metzke, T., Rogge-Solti, A., Baumgrass, A., Mendling, J., &Weske, M. (2013). Enabling Semantic Complex Event Processing in the Domain of Logistics. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *8377 LNCS*, 419–431. https://doi.org/10.1007/978-3-319-06859-6_37

Mikolov, T., Chen, K., Corrado, G., &Dean, J. (2013, January 16). Efficient estimation of word representations in vector space. *1st International Conference on Learning Representations, ICLR 2013 - Workshop Track Proceedings*. http://ronan.collobert.com/senna/

Mikolov, T., Karafiát, M., Burget, L., Honza, J., Cernock´ycernock´y, &Khudanpur, S. (2010). Recurrent neural network based language model. *INTERSPEECH-2010*, 1045–1048.

Milch, V., &Laumann, K. (2018). Sustaining safety across organizational boundaries: a qualitative study exploring how interorganizational complexity is managed on a petroleum-producing installation. *Cognition, Technology and Work*, *20*(2), 179–204.

https://doi.org/10.1007/s10111-018-0460-8

Milch, V., &Laumann, K. (2019). The influence of interorganizational factors on offshore incidents in the Norwegian petroleum industry: Challenges and future directions. *Reliability Engineering and System Safety*, *181*(October 2017), 84–96. https://doi.org/10.1016/j.ress.2018.09.002

Miles Jr, R. F. (1973). *Systems concepts: Lectures on contemporary approaches to systems*.

Morais, C., Moura, R., Beer, M., &Patelli, E. (2020). Analysis and Estimation of Human Errors From Major Accident Investigation Reports. *ASCE-ASME J Risk and Uncert in Engrg Sys Part B Mech Engrg*, *6*(1). https://doi.org/10.1115/1.4044796

Nadkarni, P. M., Ohno-Machado, L., &Chapman, W. W. (2011). Natural language processing: An introduction. In *Journal of the American Medical Informatics Association* (Vol. 18, Issue 5, pp. 544–551). Oxford University Press. https://doi.org/10.1136/amiajnl-2011-000464

Naweed, A., &Ambrosetti, A. (2015). Mentoring in the rail context: The influence of training, style, and practice. *Journal of Workplace Learning*, *27*(1), 3–18. https://doi.org/10.1108/JWL-11-2013-0098/FULL/XML

Naweed, A., Chapman, J., Vandelanotte, C., Chappel, S. E., Holtermann, A., &Straker, L. (2022). ‘Just Right’ job design: A conceptual framework for sustainable work in rail driving using the Goldilocks Work Paradigm. *Applied Ergonomics*, *105*, 103806. https://doi.org/10.1016/J.APERGO.2022.103806

Nelson, T., Ferster, C., Laberee, K., Fuller, D., &Winters, M. (2020). Crowdsourced data for bicycling research and practice. *Transport Reviews*, *41*(1), 97–114.

https://doi.org/10.1080/01441647.2020.1806943

Neto, J. L., Neto, J. L., Santos, A. D., Kaestner, C. A. A., Alexandre, N., Santos, D., A, C. A., Alex, K., Freitas, A. A., &Parana, C. (2000). *Document Clustering and Text Summarization*. http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.43.4634

Nielsen, F. Å. (2019). Ordia: A Web Application for Wikidata Lexemes. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *11762 LNCS*, 141–146. https://doi.org/10.1007/978-3-030-32327-1_28/FIGURES/3

Nielsen, L. (2011). *Rolling Stock Rescheduling in Passenger Railways Applications in Short-term Planning and in Disruption Management*. www.b-en-t.nl

Nordin, A., Nordström, G., Wilde-Larsson, B., Hallberg, A., Theander, K., Nordin, A., Nordström, G., Wilde-Larsson, B., Hallberg, A., &Theander, K. (2020). Patient Safety Culture Change over Time-Health Care Staffs' Perceptions. *Open Journal of Nursing*, *10*(3), 320–339. https://doi.org/10.4236/OJN.2020.103022

NTSB. (2014a). *Federal Aviation Administration Access to Continued Operational Safety (COS) Information During a National Transportation Safety Board Investigation*. https://www.ntsb.gov/investigations/process/Documents/NTSB-FAA-COS Letter-FinalSigned-WithLogos.pdf

NTSB. (2014b). *Guidance on Style for NTSB Written Products, 2014 and "Writing Guide" that provides guidance on the content and structure of accident reports, safety recommendations, and similar NTSB products. 2000*.

NTSB. (2016). *"What is the National Transportation Safety Board?"*

NTSB. (2020). *Preliminary Report - Railroad: Derailment of CSX Transportation Train*

K42911. https://www.ntsb.gov/investigations/AccidentReports/Pages/RRD20FR002-preliminary-report.aspx

Osei-Asamoah, A., &Lownes, N. E. (2014). Complex Network Method of Evaluating Resilience in Surface Transportation Networks. *Transportation Research Record: Journal of the Transportation Research Board*, *2467*(1), 120–128. https://doi.org/10.3141/2467-13

Oswald, M. R., &Treat, C. (2013). Identifying sea level rise vulnerability using GIS: Development of a transit inundation modeling method. *International Journal of Geoinformatics*, *9*(1), 1–10.

Ouyang, M., Hong, L., Yu, M. H., &Fei, Q. (2010). STAMP-based analysis on the railway accident and accident spreading: Taking the China-Jiaoji railway accident for example. *Safety Science*, *48*(5), 544–555. https://doi.org/10.1016/j.ssci.2010.01.002

Pagani, A., Mosquera, G., Alturki, A., Johnson, S., Jarvis, S., Wilson, A., Guo, W., &Varga, L. (2019). Resilience or robustness: Identifying topological vulnerabilities in rail networks. *Royal Society Open Science*, *6*(2). https://doi.org/10.1098/rsos.181301

Panou, K., Tzieropoulos, P., &Emery, D. (2013). Railway driver advice systems: Evaluation of methods, tools and systems. *Journal of Rail Transport Planning and Management*, *3*(4), 150–162. https://doi.org/10.1016/j.jrtpm.2013.10.005

Pant, R., Hall, J. W., &Blainey, S. P. (2016). Vulnerability assessment framework for interdependent critical infrastructures: Case-study for Great Britain's rail network. *European Journal of Transport and Infrastructure Research*, *16*(1), 174–194. https://doi.org/10.18757/ejtir.2016.16.1.3120

Parraguez, P., Eppinger, S., &Maier, A. (2016). Characterizing Design Process Interfaces as Organization Networks: Insights for Engineering Systems Management. *Systems Engineering*, *19*(2), 158–173. https://doi.org/10.1002/sys.21345

Parzen, E. (1960). *Modern probability theory and its applications*. John Wiley & Sons, Incorporated.

Patankar, M. S. (2012). *Safety culture: Building and sustaining a cultural change in aviation and healthcare*. Ashgate Publishing, Ltd..

Patankar, M. S., &Sabin, E. J. (2010). The Safety Culture Perspective. *Human Factors in Aviation*, 95–122. https://doi.org/10.1016/B978-0-12-374518-7.00004-3

Patil, M., Shinde, R. D., &Hailkar, S. S. (2008). Risk Management in Railway Projects. *International Research Journal of Engineering and Technology*, 1817. www.irjet.net

Patriarca, R., Bergström, J., DiGravio, G., &Costantino, F. (2018). Resilience engineering: Current status of the research and future challenges. In *Safety Science* (Vol. 102, pp. 79–100). Elsevier B.V. https://doi.org/10.1016/j.ssci.2017.10.005

Paul, S., Marsden, E., Verschueren, F., Tulonen, T., Ferjencik, M., Dien, Y., Simola, K., &Kopustinskas, V. (2018, October). *Accident Investigation and Learning to Improve Safety Management in Complex System: Remaining Challenges Proceedings of the 55th ESReDA Seminar Hosted by Romanian Railway Investigation Agency AGIFER*. https://doi.org/10.2760/242130

Peck, H. (2007). *Supply chain vulnerability, risk and resilience.* Global Logistic New Directions in Supply Chain Management.

Pennington, J., Socher, R., &Manning, C. D. (2014). GloVe: Global Vectors for Word Representation. In *aclweb.org*. http://nlp.

Perera, N., Dehmer, M., &Emmert-Streib, F. (2020). Named Entity Recognition and

Relation Detection for Biomedical Information Extraction. In *Frontiers in Cell and Developmental Biology* (Vol. 8). Frontiers Media S.A. https://doi.org/10.3389/fcell.2020.00673

Perrow, C. (1984). *Normal Accidents: Living with High-Risk Technologies*. New York: Basic Books.

Peters, J. L., Zevitas, C. D., Redline, S., Hastings, A., Sizov, N., Hart, J. E., Levy, J. I., Roof, C. J., &Wellenius, G. A. (2018). Aviation Noise and Cardiovascular Health in the United States: a Review of the Evidence and Recommendations for Research Direction. *Current Epidemiology Reports*, *5*(2), 140–152. https://doi.org/10.1007/S40471-018-0151-2

Peters, M. E., Ammar, W., Bhagavatula, C., &Power, R. (2017). Semi-supervised sequence tagging with bidirectional language models. *ACL 2017 - 55th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference (Long Papers)*, *1*, 1756–1765. http://arxiv.org/abs/1705.00108

Pieruschka, E. (1963). *Principles of reliability*. Prentice-Hall.

Pimm, S. L. (1984). The complexity and stability of ecosystems. *Nature*, *307*(5949), 321–326. https://doi.org/10.1038/307321a0

Piner, D., &Condry, B. (2017). International best practices in managing unplanned disruption to suburban rail services. *Transportation Research Procedia*, *25*, 4403–4410. https://doi.org/10.1016/j.trpro.2017.05.331

Pires, T., &Mosleh, A. (2011). Organizational interface failures: A historical perspective and risk analysis framework. In *Safety, Reliability and Risk Management: ESREL 2011* (pp. 218–218). https://books.google.com/books?hl=zh-

TW&lr=&id=yN3qBgAAQBAJ&oi=fnd&pg=PA218&dq=Organizational+interface+failures:+A+historical+perspective+and+risk+analysis+framework&ots=Nrym4Dpmpp&sig=tXB8xKAdlRcXI4xGrFtj562SN7g

Piscopo, A., &Simperl, E. (2019). What we talk about when we talk about Wikidata quality: A literature survey. *Proceedings of the 15th International Symposium on Open Collaboration, OpenSym 2019*. https://doi.org/10.1145/3306446.3340822

Placencia, G. (2016). Reviewing the Use of Proactive Data Analysis in Developing Rail Safety Culture. *2016 Joint Rail Conference, JRC 2016*. https://doi.org/10.1115/JRC2016-5799

Poirier, C., Adelé, S., &Burkhardt, J. (2020). Individual accidents at the interface between platform, train and tracks (PT 2 I) in the subway: a literature review. *Cognition, Technology & Work*, *23*(2), 203–224. https://doi.org/10.1007/s10111-020-00624-1

Pramanik, A., Sarkar, S., Sai Siddharth, V., Maiti, J., &Mitra, P. (2021). Semi-automated Ontology Creation and Upgradation for Rail-Road Incidents: A Case of a Steel Plant in India. *Lecture Notes in Networks and Systems*, *164*, 285–294. https://doi.org/10.1007/978-981-15-9774-9_28

Priestley, K., &Lee, G. (2008). Human factors in railway operations. *2008 International Conference on Railway Engineering - Challenges for Railway Transportation in Information Age*, 1–5. https://ieeexplore.ieee.org/abstract/document/4730852

Putnam, R. D. (2000). *Bowling Alone*. http://bowlingalone.com/socialcapital.php3/

Qurashi, A. W., Holmes, V., &Johnson, A. P. (2020). Document Processing: Methods for Semantic Text Similarity Analysis. *INISTA 2020 - 2020 International Conference on INnovations in Intelligent SysTems and Applications, Proceedings*.

https://doi.org/10.1109/INISTA49547.2020.9194665

RAIB. (2005). *Guidance on the Railways (Accident Investigation and Reporting)*. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/456936/guidance_to_rair_regs_v4.pdf

RAIB. (2008). *Passenger trapped in a closed train door, Tooting Broadway, Northern Line, London Underground*. www.raib.gov.uk.

RAIB. (2011). *Investigation into the safety of automatic open level crossings on Network Rail's managed infrastructure*. https://assets.publishing.service.gov.uk/media/547c8fda40f0b60241000167/R122011_110728_AOCLs_Class_Inv.pdf

RAIB. (2019). *An introduction to the RAIB*. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/829080/190118_Leaflet_01_Introduction.pdf

RAIB. (2020). *Signal passed at stop and near miss, Deansgate-Castlefield tram stop, Manchester*. www.gov.uk/raib

Railtrack, C. E., &House, R. (1997). Railway Safety Cases. *Safety and Reliability of Software Based Systems*, 317–322. https://doi.org/10.1007/978-1-4471-0921-1_18

Rao, S. (2007). Safety culture and accident analysis—A socio-management approach based on organizational safety social capital. *Journal of Hazardous Materials*, *142*(3), 730–740. https://doi.org/10.1016/J.JHAZMAT.2006.06.086

Raslear, T. (2006). A Pilot Examination of a Joint Railroad Management-Labor Approach to Root Cause Analysis of Accidents, Incidents, and Close Calls in a Diesel and Car Repair Shop Environment. *Research Results*, *November*.

Rasmussen, J. (1997). Risk management in a dynamic society: a modelling problem. *Safety Science*, *27*(2–3), 183–213. https://doi.org/10.1016/S0925-7535(97)00052-0

Rasmussen, J., &Svedung, I. (2000). *Proactive Risk Management in a Dynamic Society-Swedish Rescue Services Agency.* Swedish Rescue Services Agency.

Rausand, M. (2013). *Risk assessment: theory, methods, and applications* (John Wiley & Sons. (ed.); Vol. 115).

Read, G. J. M., Cox, J. A., Hulme, A., Naweed, A., &Salmon, P. M. (2021). What factors influence risk at rail level crossings? A systematic review and synthesis of findings using systems thinking. *Safety Science*, *138*, 105207. https://doi.org/10.1016/J.SSCI.2021.105207

Read, G. J. M., Salmon, P. M., &Lenné, M. G. (2013). Sounding the warning bells: The need for a systems approach to understanding behaviour at rail level crossings. *Applied Ergonomics*, *44*(5), 764–774. https://doi.org/10.1016/J.APERGO.2013.01.007

Read, G. J., Shorrock, S., Walker, G. H., & Salmon, P. M. (2021). State of science: Evolving perspectives on 'human error'. Ergonomics, 64(9), 1091-1114.

Reason, J. (1990). *Human error*. https://books.google.com/books?hl=zh-TW&lr=&id=WJL8NZc8lZ8C&oi=fnd&pg=PR9&ots=AnMk7gdh29&sig=-8vOkqCDeIQEawd3FnEKvsd-d3s

Reason, J. (2000). *Human error: models and management*. https://www.bmj.com/content/320/7237/768?ath_user=efschouston002&ath_ttok=%253CTPEEDaMygDfSrj8WwQ%253E

Reggiani, A., Nijkamp, P., &Lanzi, D. (2015). Transport resilience and vulnerability: The role of connectivity. *Transportation Research Part A: Policy and Practice*, *81*, 4–15.

https://doi.org/10.1016/j.tra.2014.12.012

Reimers, N., &Gurevych, I. (2019). Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. *EMNLP-IJCNLP 2019 - 2019 Conference on Empirical Methods in Natural Language Processing and 9th International Joint Conference on Natural Language Processing, Proceedings of the Conference*, 3982–3992. https://doi.org/10.48550/arxiv.1908.10084

Restrepo, C. E., Simonoff, J. S., &Zimmerman, R. (2009). Causes, cost consequences, and risk implications of accidents in US hazardous liquid pipeline infrastructure. *International Journal of Critical Infrastructure Protection*, *2*(1–2), 38–50. https://doi.org/10.1016/j.ijcip.2008.09.001

Reyes-Peña, C., &Tovar-Vidal, M. (2019). Ontology: Components and Evaluation, a Review. *Research in Computing Science*, *148*(3), 257–265. https://doi.org/10.13053/rcs-148-3-21

Roberts, M. E., Stewart, B. M., &Tingley, D. (2019). stm: An R Package for Structural Topic Models. *Journal of Statistical Software*, *91*, 1–40. https://doi.org/10.18637/JSS.V091.I02

Roberts, M. E., Stewart, B. M., Tingley, D., &Airoldi, E. M. (2013). The structural topic model and applied social science. *Adv. Neural Inf. Process. Syst. Workshop Top. Models: Comput. Appl. Eval.*, 1–20.

Roque, C., Lourenço Cardoso, J., Connell, T., Schermers, G., &Weber, R. (2019). Topic analysis of Road safety inspections using latent dirichlet allocation: A case study of roadside safety in Irish main roads. *Accident Analysis & Prevention*, *131*, 336–349. https://doi.org/10.1016/J.AAP.2019.07.021

Rosoff, H., &vonWinterfeldt, D. (2007). A Risk and Economic Analysis of Dirty Bomb Attacks on the Ports of Los Angeles and Long Beach. *Risk Analysis*, *27*(3), 533–546. https://doi.org/10.1111/j.1539-6924.2007.00908.x

Sakor, A., Singh, K., Patel, A., &Vidal, M. E. (2020). Falcon 2.0: An Entity and Relation Linking Tool over Wikidata. *International Conference on Information and Knowledge Management, Proceedings*, 3141–3148. https://doi.org/10.1145/3340531.3412777

Salmon, P. M., Lenné, M. G., Read, G. J. M., Mulvihill, C. M., Cornelissen, M., Walker, G. H., Young, K. L., Stevens, N., &Stanton, N. A. (2016). More than meets the eye: Using cognitive work analysis to identify design requirements for future rail level crossing systems. *Applied Ergonomics*, *53*, 312–322. https://doi.org/10.1016/J.APERGO.2015.06.021

Salmon, P. M., Read, G. J. M., Stanton, N. A., &Lenné, M. G. (2013). The crash at Kerang: Investigating systemic and psychological factors leading to unintentional non-compliance at rail level crossings. *Accident Analysis and Prevention*, *50*, 1278–1288. https://doi.org/10.1016/j.aap.2012.09.029

Salzberg, S. L. (1994). C4.5: Programs for Machine Learning by J. Ross Quinlan. Morgan Kaufmann Publishers, Inc., 1993. *Machine Learning*, *16*(3), 235–240. https://doi.org/10.1007/bf00993309

Sammarco, J. J. (2003). *A normal accident theory-based complexity assessment methodology for safety-related embedded computer systems.* West Virginia University.

Santos-Reyes, J., &Beard, A. N. (2006). A Systemic Analysis of the Paddington Railway Accident. *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, *220*(2), 121–151. https://doi.org/10.1243/09544097JRRT33

Santos-Reyes, J., &Beard, A. N. (2009). A systemic analysis of the Edge Hill railway accident. *Accident Analysis and Prevention*, *41*(6), 1133–1144. https://doi.org/10.1016/j.aap.2008.05.004

Sanusi, K., Ismail, S., &Shariff, N. M. M. (2015). Customized safety behavior checklist for scrubber installation at construction site. *Proceedings of the 26th International Business Information Management Association Conference - Innovation Management and Sustainable Economic Competitive Advantage: From Regional Development to Global Growth, IBIMA 2015*, 505–514.

Sarewitz, D., Pielke, R., &Keykhah, M. (2003). Vulnerability and risk: Some thoughts from a political and policy perspective. In *Risk Analysis* (Vol. 23, Issue 4, pp. 805–810). John Wiley & Sons, Ltd. https://doi.org/10.1111/1539-6924.00357

Schein, E. H. (1992). *Organizational Culture and Leadership* (2nd Editio). Jossey-Bass, San Francisco.

Settles, B. (2003). Biomedical Named Entity Recognition Using Conditional Random Fields and Rich Feature Sets. *Proceedings of the International Joint Workshop on Natural Language Processing in Biomedicine and Its Applications (NLPBA/BioNLP)*, 107–110. http://us.expasy.org/sprot/

Shappell, S., &Wiegmann, D. (2000). The Human Factors Analysis and Classification System--HFACS. In *Publications*. https://commons.erau.edu/publication/737

Shenoy, K., Ilievski, F., Garijo, D., Schwabe, D., &Szekely, P. (2022). A study of the quality of Wikidata. *Journal of Web Semantics*, *72*, 100679. https://doi.org/10.1016/J.WEBSEM.2021.100679

Shokri, S., Safa, M., Haas, C. T., Haas, R. C. G., Maloney, K., &MacGillivray, S. (2012).

Interface Management Model for Mega Capital Projects. *Construction Research Congress 2012*, 447–456. https://doi.org/10.1061/9780784412329.045

Shrivastava, S., Sonpar, K., &Pazzaglia, F. (2009). Normal Accident Theory versus High Reliability Theory: A resolution and call for an open systems view of accidents. *Human Relations*, *62*(9), 1357–1390. https://doi.org/10.1177/0018726709339117

Shu, C., Dosyn, D., Lytvyn, V., Vysotska, V., Sachenko, A., &Jun, S. (2019). Building of the Predicate Recognition System for the NLP Ontology Learning Module. *Proceedings of the 2019 10th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications, IDAACS 2019*, *2*, 802–808. https://doi.org/10.1109/IDAACS.2019.8924410

Silva, F., Lambe, T. W., &Marr, W. A. (2008). Probability and Risk of Slope Failure. *Journal of Geotechnical and Geoenvironmental Engineering*, *134*(12), 1691–1699. https://doi.org/10.1061/(asce)1090-0241(2008)134:12(1691)

Sintek, M., &Decker, S. (2002). TRIPLE-A query, inference, and transformation language for the Semantic Web. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *2342 LNCS*, 364–378. https://doi.org/10.1007/3-540-48005-6_28/COVER

Sizov, G., &Öztürk, P. (2013). *Automatic Extraction of Reasoning Chains from Textual Reports*. 61–69. https://aclanthology.org/W13-5009

Sklet, S. (2002). *Methods for accident investigation (ROSS (NTNU) 200208).* . NTNU. https://scholar.google.com/scholar?hl=zh-TW&as_sdt=0%2C5&q=Sklet%2C+S.+%282002%29.+Methods+for+Accident+Investigation.+ROSS+report+200208.&btnG=

Soleimani, S., Mohammadi, A., Chen, J., &Leitner, M. (2019). Mining the highway-rail grade crossing crash data: A text mining approach. *Proceedings - 18th IEEE International Conference on Machine Learning and Applications, ICMLA 2019*, 1063–1068. https://doi.org/10.1109/ICMLA.2019.00179

Song, T., Zhong, D., &Zhong, H. (2012). A STAMP analysis on the China-Yongwen railway accident. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *7612 LNCS*, 376–387. https://doi.org/10.1007/978-3-642-33678-2_32

Srathongkhruen, S., &Fraszczyk, A. (2021). Safety Culture in a Railway Maintenance Environment: A Case Study of Bangkok Metro Network. In *Lecture Notes in Mobility*. https://doi.org/10.1007/978-981-15-5979-2_3

Stanton, N. A., Salmon, P. M., Walker, G. H., &Stanton, M. (2019). Models and methods for collision analysis: A comparison study based on the Uber collision with a pedestrian. *SAFETY SCIENCE*, *120*, 117–128. https://doi.org/10.1016/j.ssci.2019.06.008

Størseth, F., &Tinmannsvik, R. K. (2012). The critical re-action: Learning from accidents. *Safety Science*, *50*(10), 1977–1982. https://doi.org/10.1016/J.SSCI.2011.11.003

Sun, Huang, Y., Chen, Y., &Yao, L. (2018). Vulnerability assessment of urban rail transit based on multi-static weighted method in Beijing, China. *Transportation Research Part A: Policy and Practice*, *108*, 12–24. https://doi.org/10.1016/j.tra.2017.12.008

Syeda, K. N., Shirazi, S. N., Asad, S., Naqvi, A., &Parkinson, H. J. (2017). Exploiting Natural Language Processing for Analysing Railway Incident Reports. *IGI Global Publishing*, 1–18. http://www.digitalrail.co.uk/wp-content/uploads/2018/01/Exploiting-Natural-Language-Processing-for-Analysing-Railway-Incident-Reports.pdf

Syeda, K. N., Shirazi, S. N., Naqvi, S. A. A., Parkinson, H. J., &Bamford, G. (2019). Big Data and natural language processing for analysing railway safety: Analysis of railway incident reports. *Human Performance Technology: Concepts, Methodologies, Tools, and Applications.*, 781–809. http://search.ebscohost.com.proxy-ub.rug.nl/login.aspx?direct=true&db=psyh&AN=2019-25313-040&site=ehost-live&scope=site

Szymula, C., &Bešinović, N. (2020). Passenger-centered vulnerability assessment of railway networks. *Transportation Research Part B: Methodological*, *136*, 30–61. https://doi.org/10.1016/j.trb.2020.03.008

Tang, J., Liu, X., &Wang, W. (2022). Causations Analysis of a High-Speed Railway Accident Using Extended 24Model with Interval Type-2 Fuzzy DEMATEL Method. *Https://Doi.Org/10.1080/01969722.2022.2146926*. https://doi.org/10.1080/01969722.2022.2146926

Taylor, M. A. P., &D'Este, G. M. (2007). Transport network vulnerability: A method for diagnosis of critical locations in transport infrastructure systems. In *Advances in Spatial Science* (Issue 9783540680550, pp. 9–30). Springer International Publishing. https://doi.org/10.1007/978-3-540-68056-7_2

Taysom, E., &Crilly, N. (2017). Resilience in Sociotechnical Systems: The Perspectives of Multiple Stakeholders. *She Ji*, *3*(3), 165–182. https://doi.org/10.1016/j.sheji.2017.10.011

Teperi, A. M., Paajanen, T., Asikainen, I., &Lantto, E. (2023). From must to mindset: Outcomes of human factor practices in aviation and railway companies. *Safety Science*, *158*, 105968. https://doi.org/10.1016/J.SSCI.2022.105968

Tharani, K. (2021). Much more than a mere technology: A systematic review of Wikidata in libraries. *The Journal of Academic Librarianship*, *47*(2), 102326. https://doi.org/10.1016/J.ACALIB.2021.102326

Thatcher, A., Nayak, R., &Waterson, P. (2019). Human factors and ergonomics systems-based tools for understanding and addressing global problems of the twenty-first century. *Ergonomics*, *63*(3), 367–387. https://doi.org/10.1080/00140139.2019.1646925

Tretten, P., &Candell, O. (2021). AHFE (7) - Digitalization of Railway Maintenance: A Situation Awareness Perspective. In *Advances in Human Factors and System Interactions: Vol. NA* (Issue NA). https://doi.org/10.1007/978-3-030-79816-1_25

TSB. (1990). *Rail transportation safety recommendations - Transportation Safety Board of Canada*. https://www.tsb.gc.ca/eng/recommandations-recommendations/rail/index.html

TSB. (2014a). *“About the TSB.”*

TSB. (2014b). *Railway Investigation Report R13D0054*. https://www.tsb.gc.ca/eng/rapports-reports/rail/2013/R13D0054/R13D0054.html

TSB. (2019). *Railway Investigation Report R18W0133*. https://www.tsb.gc.ca/eng/rapports-reports/rail/2018/r18w0133/r18w0133.html

TSB. (2022). *Policy on Occurrence Classification - Transportation Safety Board of Canada*. https://www.tsb.gc.ca/eng/lois-acts/evenements-occurrences.html

Turki, H., Shafee, T., Hadj Taieb, M. A., BenAouicha, M., Vrandečić, D., Das, D., &Hamdi, H. (2019). Wikidata: A large-scale collaborative ontological medical database. *Journal of Biomedical Informatics*, *99*, 103292. https://doi.org/10.1016/J.JBI.2019.103292

Turner, B. A., &Pidgeon, N. F. (1997). *Man-made disasters.* https://scholar.google.com/scholar?hl=zh-TW&as_sdt=0%2C5&q=Turner+and+Pidgeon+%281997%29&btnG=

Turnquist, M., &Vugrin, E. (2013). Design for resilience in infrastructure distribution networks. *Environmentalist*, *33*(1), 104–120. https://doi.org/10.1007/s10669-012-9428-z

Tutcher, J. (2015). Ontology-driven data integration for railway asset monitoring applications. *Proceedings - 2014 IEEE International Conference on Big Data, IEEE Big Data 2014*, 85–95. https://doi.org/10.1109/BigData.2014.7004436

Underwood, P., &Waterson, P. (2013). Systemic accident analysis: Examining the gap between research and practice. *Accident Analysis and Prevention*, *55*, 154–164. https://doi.org/10.1016/j.aap.2013.02.041

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., &Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, *2017-December*, 5999–6009. https://arxiv.org/abs/1706.03762v5

Vaughan, D. (1999). THE DARK SIDE OF ORGANIZATIONS: Mistake, Misconduct, and Disaster. *Annual Review of Sociology*, *25*(1), 271–305. https://doi.org/10.1146/annurev.soc.25.1.271

Ventroux, J., Marle, F., &Vidal, L.-A. (2017). *Optimize The Supervision Of Complex Projects By Taking Into Account Interactions Between Actors*. https://hal.archives-ouvertes.fr/hal-01589602

Villalba Sanchis, I., Insa Franco, R., Martínez Fernández, P., Salvador Zuriaga, P., &Font Torres, J. B. (2020). Risk of increasing temperature due to climate change on high-speed rail network in Spain. *Transportation Research Part D: Transport and Environment*, *82*. https://doi.org/10.1016/j.trd.2020.102312

Waagmeester, A., Stupp, G., Burgstaller-Muehlbacher, S., Good, B. M., Griffith, M., Griffith,

O. L., Hanspers, K., Hermjakob, H., Hudson, T. S., Hybiske, K., Keating, S. M., Manske, M., Mayers, M., Mietchen, D., Mitraka, E., Pico, A. R., Putman, T., Riutta, A., Queralt-Rosinach, N., …Su, A. I. (2020). Wikidata as a knowledge graph for the life sciences. *ELife*, *9*. https://doi.org/10.7554/ELIFE.52614

Waagmeester, A., Willighagen, E. L., Su, A. I., Kutmon, M., Gayo, J. E. L., Fernández-Álvarez, D., Groom, Q., Schaap, P. J., Verhagen, L. M., &Koehorst, J. J. (2021). A protocol for adding knowledge to Wikidata: aligning resources on human coronaviruses. *BMC Biology 2021 19:1*, *19*(1), 1–14. https://doi.org/10.1186/S12915-020-00940-Y

Wang, C.-H., &Liu, Y.-J. (2012). Omnidirectional safety culture analysis and discussion for railway industry. *Safety Science*, *50*(5), 1196–1204. https://doi.org/10.1016/j.ssci.2011.12.031

Wang, J., Muddada, R. R., Wang, H., Ding, J., Lin, Y., Liu, C., &Zhang, W. (2016). Toward a resilient holistic supply chain network system: Concept, review and future direction. In *IEEE Systems Journal* (Vol. 10, Issue 2, pp. 410–421). Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/JSYST.2014.2363161

Wang, S. S., Dong, H. H., Zhou, Y., Jia, L. M., Qin, Y., &IEEE. (2017). Exploring Traffic Accident Locations from Natural Language Based on Spatial Information Retrieval. In *2017 29TH CHINESE CONTROL AND DECISION CONFERENCE (CCDC)* (Issue 29th Chinese Control And Decision Conference (CCDC), pp. 3490–3495).

Wang, X., Koç, Y., Derrible, S., Ahmad, S. N., Pino, W. J. A., &Kooij, R. E. (2017). Multi-criteria robustness analysis of metro networks. *Physica A: Statistical Mechanics and Its Applications*, *474*, 19–31. https://doi.org/10.1016/j.physa.2017.01.072

Wheway, J. L., &Jun, G. T. (2021). Adopting systems models for multiple incident analysis:

utility and usability. *International Journal for Quality in Health Care : Journal of the International Society for Quality in Health Care*, *33*(4), NA-NA. https://doi.org/10.1093/intqhc/mzab135

Woods, D. D., & Cook, R. I. (2003). Mistaking error. Patient safety handbook, 95, 108.

Systems Theory, Sourcebook of Family Theories and Methods 325. https://doi.org/10.1007/978-0-387-85764-0_14

Whittingham, R. (2004). *The blame machine: Why human error causes accidents.* https://scholar.google.com/scholar?hl=zh-TW&as_sdt=0%2C5&q=human+error++railway+accident&btnG=#d=gs_cit&u=%2Fscholar%3Fq%3Dinfo%3A4ccyH4iAqAoJ%3Ascholar.google.com%2F%26output%3Dcite%26scirp%3D0%26hl%3Dzh-TW

Whittingham, R. B. (2012). Preventing corporate accidents: An ethical approach. In *Preventing Corporate Accidents: An Ethical Approach*. https://doi.org/10.4324/9780080570297

Wiegmann, D., &Shappell, S. (2017). *A human error approach to aviation accident analysis: The human factors analysis and classification system*. https://books.google.com/books?hl=zh-TW&lr=&id=jitEDwAAQBAJ&oi=fnd&pg=PT8&ots=PEUtxjKzwm&sig=FdZupxYLiaeff6W-1W0dH3DtNAg

Wilson, J. R., &Norris, B. J. (2005). Rail human factors: Past, present and future. *Applied Ergonomics*, *36*(6), 649–660. https://doi.org/10.1016/J.APERGO.2005.07.001

Wittmer, A., &Mark, R. (2021). Human Factors in Aviation. In *Classroom Companion: Business: Vol. NA* (Issue NA, pp. 355–386). https://doi.org/10.1007/978-3-030-79549-

8_12

Worton, K. E. (2012). Using socio-technical and resilience frameworks to anticipate threat. *Proceedings - 2nd Workshop on Socio-Technical Aspects in Security and Trust, STAST 2012, Co-Located with 25th IEEE Computer Security Foundations Symposium, CSF 2012*, 19–26. https://doi.org/10.1109/STAST.2012.16

Wrigstad, J., Bergström, J., &Gustafson, P. (2014). Mind the gap between recommendation and implementation - Principles and lessons in the aftermath of incident investigations: A semi-quantitative and qualitative study of factors leading to the successful implementation of recommendations. *BMJ Open*, *4*(5), 5326. https://doi.org/10.1136/bmjopen-2014-005326

Wu, C., Wu, F., Qi, T., &Huang, Y. (2020). Named Entity Recognition with Context-Aware Dictionary Knowledge. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *12522 LNAI*, 129–143. https://doi.org/10.1007/978-3-030-63031-7_10

Wu, H., Zhong, B., Medjdoub, B., Xing, X., &Jiao, L. (2020). An ontological metro accident case retrieval using CBR and NLP. *Applied Sciences (Switzerland)*, *10*(15). https://doi.org/10.3390/APP10155298

Wu, Y., Schuster, M., Chen, Z., Le, Q.V., Norouzi, M., Macherey, W., Krikun, M., Cao, Y., Gao, Q., Macherey, K., Klingner, J., Shah, A., Johnson, M., Liu, X., Kaiser, Ł., Gouws, S., Kato, Y., Kudo, T., Kazawa, H., …Dean, J. (2016). *Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation*. http://arxiv.org/abs/1609.08144

Wullems, C., Toft, Y., &Dell, G. (2013). Improving level crossing safety through enhanced data recording and reporting: The CRC for rail innovation's baseline rail level crossing video project. *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, *227*(5), 554–559. https://doi.org/10.1177/0954409713502014

Xia, M., Li, X., Jiang, F., &Wang, S. (2012). Cause analysis and countermeasures of locomotive runaway accident based on fault tree analysis method. *Procedia Engineering*, *45*, 38–42. https://doi.org/10.1016/j.proeng.2012.08.117

Xie, Q., Dai, Z., Hovy, E., Luong, M.-T., &Le, Q.V. (2019). Unsupervised Data Augmentation for Consistency Training. *ArXiv*. http://arxiv.org/abs/1904.12848

Xing, F. Z., Cambria, E., &Welsch, R. E. (2018). Natural language based financial forecasting: a survey. *Artificial Intelligence Review*, *50*(1), 49–73. https://doi.org/10.1007/s10462-017-9588-9

Yamashiro, H., &Nonaka, H. (2019). *Development of an Unsupervised Learning Methods for Classification of Accident Reports without Code Information*. 867–873.

Yang, H., Lo, H., &Tang, W. H. (2000). Travel time versus capacity reliability of a road network. *Reliability of Transport Networks*, 119–138.

Yang, J., &Anwar, A. M. (2016). Social Media Analysis on Evaluating Organisational Performance: A Railway Service Management Context. *Proceedings - 2016 IEEE 14th International Conference on Dependable, Autonomic and Secure Computing, DASC 2016, 2016 IEEE 14th International Conference on Pervasive Intelligence and Computing, PICom 2016, 2016 IEEE 2nd International Conference on Big Data*, 835–841. https://doi.org/10.1109/DASC-PICom-DataCom-CyberSciTec.2016.143

Yang, J., Han, S. C., &Poon, J. (2021). *A Survey on Extraction of Causal Relations from Natural Language Text*. https://arxiv.org/abs/2101.06426v1

Yang, Y., UY, M. C. S., &Huang, A. (2020). FinBERT: A Pretrained Language Model for

Financial Communications. *ArXiv*. http://arxiv.org/abs/2006.08097

Yap, M., &Cats, O. (2019, June 1). Analysis and Prediction of Disruptions in Metro Networks. *MT-ITS 2019 - 6th International Conference on Models and Technologies for Intelligent Transportation Systems*. https://doi.org/10.1109/MTITS.2019.8883320

Yildirim, S., Jothimani, D., Kavaklioglu, C., &Başar, A. (2019). Classification of »hot News» for Financial Forecast Using NLP Techniques. *Proceedings - 2018 IEEE International Conference on Big Data, Big Data 2018*, 4719–4722. https://doi.org/10.1109/BigData.2018.8621903

Younes, N. (2005). Breaking the safety plateau: A successful implementation of behaviour-based safety across a range of cultures, facilities, and functions. *2005 SPE Asia Pacific Health, Safety and Environment Conference and Exhibition - Proceedings*, *2005*, 67–70. https://doi.org/10.2118/95366-ms

Young, M. S., Stanton, N. A., &Walker, G. H. (2006). In loco intellegentia: human factors for the future European train driver. *International Journal of Industrial and Systems Engineering*, *1*(4), 485–501. https://doi.org/10.1504/IJISE.2006.010388

Young, T., Hazarika, D., Poria, S., &Cambria, E. (2018). Recent trends in deep learning based natural language processing [Review Article]. *IEEE Computational Intelligence Magazine*, *13*(3), 55–75. https://doi.org/10.1109/MCI.2018.2840738

Yu, G., Zheng, W., Wang, L., &Zhang, Z. (2018). Identification of Significant Factors Contributing to Multi-attribute Railway Accidents Dataset (MARA-D) Using SOM Data Mining. *IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC*, *2018-Novem*, 170–175. https://doi.org/10.1109/ITSC.2018.8569336

Yu, J., Bohnet, B., &Poesio, M. (2019). Neural Mention Detection. *LREC 2020 - 12th*

International Conference on Language Resources and Evaluation, Conference Proceedings, 1–10. https://doi.org/10.48550/arxiv.1907.12524

Yuyua, G., Xu, K., Li, J., Yao, X.-W., Wu, C.-Y., Li, S.-Y., Yan, F., Zhang, J., Xu, Q.-W., &Ji Zhang. (2021). Interaction Theory of Hazard-Target System. In *NA: Vol. NA* (Issue NA, p. NA-NA). https://doi.org/NA

Zhan, Q., &Zheng, W. (2016). An in-depth learn from incidents (LFI) approach based on HFACS-RAs. *2016 IEEE International Conference on Intelligent Rail Transportation, ICIRT 2016*, 177–185. https://doi.org/10.1109/ICIRT.2016.7588729

Zhang, D. ming, Du, F., Huang, H., Zhang, F., Ayyub, B. M., &Beer, M. (2018). Resiliency assessment of urban rail transit networks: Shanghai metro as an example. *Safety Science*, *106*, 230–243. https://doi.org/10.1016/j.ssci.2018.03.023

Zhang, W., Chen, J., Paudel, B., Zhu, H., Wang, L., Zhang, W., Bernstein, A., &Chen, H. (2019). Iteratively learning embeddings and rules for knowledge graph reasoning. *The Web Conference 2019 - Proceedings of the World Wide Web Conference, WWW 2019*, 2366–2377. https://doi.org/10.1145/3308558.3313612

Zhang, X. (2014). Chengqing Zong: Statistical natural language processing (second edition). *Machine Translation*, *28*(2), 155–158. https://doi.org/10.1007/s10590-014-9150-3

Zhang, X., &Miller-Hooks, E. (2015). Scheduling Short-Term Recovery Activities to Maximize Transportation Network Resilience. *Journal of Computing in Civil Engineering*, *29*(6), 04014087. https://doi.org/10.1061/(asce)cp.1943-5487.0000417

Zhang, Y., Xu, J., Chen, H., Wang, J., Wu, Y., Prakasam, M., &Xu, H. (2016). Chemical named entity recognition in patents by domain knowledge and unsupervised feature learning. *Database*, *2016*. https://doi.org/10.1093/database/baw049

Zhao, R., Ma, X., Dong, H., Yan, H., Chen, F., &Jia, L. (2022). *Data-Driven Ontology Construction Method for Railway Derailment Accidents*. 524–533. https://doi.org/10.1007/978-981-16-9913-9_59

Zhou, J. L., &Lei, Y. (2018). Paths between latent and active errors: Analysis of 407 railway accidents/incidents' causes in China. *Safety Science*, *110*(November 2017), 47–58. https://doi.org/10.1016/j.ssci.2017.12.027

Zhou, Y., Wang, J., &Yang, H. (2019). Resilience of Transportation Systems: Concepts and Comprehensive Review. *IEEE Transactions on Intelligent Transportation Systems*, *20*(12), 4262–4276. https://doi.org/10.1109/TITS.2018.2883766