CHEMICAL ENGINEERING TRANSACTIONS
VOL. 111, 2024
A publication of AIDIC
The Italian Association of Chemical Engineering
Online at www.cetjournal.it
Guest Editors: Valerio Cozzani, Bruno Fabiano, Genserik Reniers

ISBN 979-12-81206-11-3; **ISSN** 2283-9216
DOI: 10.3303/CET24111086

Towards Operational Resilience Supported by Artificial Intelligence for Cargo Handling and Container Activities

Margherita Pettinato, Evgeniia Taubert, Tomaso Vairo, Fabio Currò, Bruno Fabiano*

DICCA Civil, Chemical and Environmental Engineering Dept., Genova University, Via Opera Pia, 15 -16145 Genova, Italy
brown@unige.it

Resilience engineering focuses on enhancing the ability of complex systems, organisations and processes to adapt to and recover from disruptions and unexpected events while maintaining their essential functions. It relies on the detection of early precursor of system failure states, on flexibility and controllability of processes, protective barriers construction and minimization of recovery time. Additionally, each process of the system should be characterised by learnability due to constant feedback and competently built management. It has been widely recognized the pivotal role of AI algorithms, which can analyse big data collected from sensors, historical records, and external sources to identify patterns, detect anomalies, and make predictive assessments. This paper critically explores the possibilities of applying text mining and Natural Language Processing techniques for entity extraction to construct an organisational resilience model more efficiently. Accordingly, visualization techniques are used to understand data patterns and trends and identify any areas for improvement (EDA – Exploratory Data Analysis). The textual analyses were based on accident reports obtained from Genoa port companies over 10 years. The data-driven decision-making enables proactive risk mitigation, early identification of potential failures, and optimization of safety protocols, and, in perspective, optimising the learning capacity of the port resilience system.

1. Introduction

A common property of many complex systems is resilience, which is the ability of a system to respond to perturbations, internal failures and environmental events by absorbing disturbances and/or reorganising to preserve its functions. In other words, resilience is the defence function of a system that includes avoidance, absorption, adaptation, and recovery from disruptions (Dinh et al., 2012). Currently, resilience is an interdisciplinary concern that involves fields such as natural, social, and engineering sciences. Resilience engineering focuses on understanding and enhancing the ability of complex systems, organisations, and processes to adapt and recover from failures and unforeseen events, while maintaining their core functions (Fraccascia et al., 2018). In particular, the ability to manage and coordinate the activities in an agile, flexible, and fast-recovery way was recognized as the key factor in addressing organizational unexpected events towards a resilient response (Fabiano et al., 2024). In the peculiar industrial port environment, a dynamic resilience model involving the probability of operational perturbations and their updates based on the hard (failures) and soft (process variables deviations) was recently proposed (Vairo et al., 2021). By analysing the interplay between organizational and technical factors and their role on preventing and mitigating barriers, it can be at identifying critical areas needing improvement (Vairo et al., 2023). Thus, the main objectives of resilience management are early detection of failures at the normal state, the reduction of the failure probability, mitigation of the potential consequences, minimization of recovery and regeneration time after failure. The resilience of any process at the micro and macro level ensures the system if preventive barriers fail, also due to the ability to learn that a resilient system is expected to show, due to constant feedback from the system and properly constructed management, as shown in Figure 1. While the specific quantitative parameters of resilience may vary depending on the context and objectives of the system, some general categories can be identified to assess the likelihood of transition from 'normal operation' to 'near misses', 'failure' or 'catastrophic failure development'.

Paper Received: 30 January 2024; Revised: 23 June 2024; Accepted: 11 July 2024
Please cite this article as: Pettinato M., Taubert E., Vairo T., Currò F., Fabiano B., 2024, Towards Operational Resilience Supported by Artificial Intelligence for Cargo Handling and Container Activities, Chemical Engineering Transactions, 111, 511-516 DOI:10.3303/CET24111086

Table 1 presents a novel description of critical components of resilience, which became an interpretation of the four resilience aspects mentioned by Bragatto et al. (2021), moving towards the resilience of a live organism.

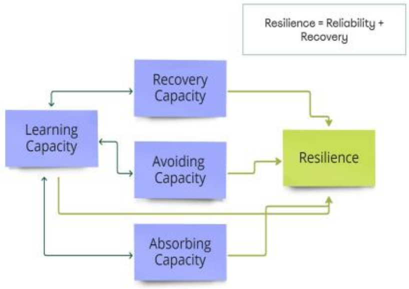

Figure 1: Relationship of the main components of resilience.

Table 1: Critical resilience components.

Avoiding capacity	Absorption capacity	Recovery capacity	Learning capacity
Ability to assess risks and develop strategies to mitigate and avoid hazards. Identifying the most vulnerable steps/components of the process and implementation of protective and preventive barriers. *No occurrence of unwanted event.*	Ability to maintain output through adaptability and flexibility of production processes. *Unwanted event occurrence: failure of one of the protective layers.*	The ability to return to normal performance as quickly as possible. *Recovery from an unwanted event.*	The ability of the system to share information between its components and learn from mistakes. Revision of the risk mitigation strategy.

The focus of this paper was shifted towards establishing a link between the information-sharing capacity (learning capacity) and hazard assessment (anticipation capacity) of the system. Most of the information which is circulating in any company and industrial environment is provided in unstructured textual form. A reliable source of information for hazard analysis and risk assessment is Periodic Health Assessment (PHA) documentation and accident reports. Information permits to extract hazards, their causes, and consequences from the text (entity extraction), allowing one to find the relationship between risks and anomalies to predict accidents (accident prediction). This facet is particularly relevant for management errors, control systems failures, or human anomalies detection. Text Mining (TM) and Natural Language Processing (NLP) are solutions to simplify the analysis of large amounts of unstructured textual data, extract useful features, and understand the context of natural language by machine learning algorithms. This article presents a textual analysis of accident descriptions which were obtained from Genoa port container companies for the period 2012-2023. A sequence of text processing procedures common for TM and NLP allowed to visualise the relationship between accidents, operations performed and injury severity. For retaining contextual information and a more detailed analysis of possible causes and vulnerable system components, a co-occurrence network was also applied to identify any areas for improvement (EDA – Exploratory Data Analysis).

2. Methodology

TM is the process of automatically extracting valuable information, patterns, and relationships from textual data. NLP, in turn, allows by using techniques of computational linguistics, statistics, and machine learning (including Deep Learning) to translate natural language into a numerical format while preserving semantic meaning and context for further computer interpretation and replicability. TM is closely related to NLP as it utilises similar approaches at least in the initial stage. One of these approaches is the Term Frequency-Inverse Document Frequency (TF-IDF). TF-IDF is a statistical model commonly used in traditional machine learning algorithms for text analysis. It represents the importance of a word in a document (a sentence) relative to a collection of documents (a set of sentences in a corpus of text or multiple texts), generating a term-document matrix of TF-IDF estimates. Thus, higher weights are assigned to words that do not occur frequently in the dataset but have high frequency in several documents. This method is limited in capturing semantic and contextual information of the text, which is well handled by methods such as Word2Vec and Bidirectional Encoder Representations from Transformers (BERT). Word2Vec is a word embedding model, which captures semantic and contextual information of words in a continuous vector space. It is an unsupervised learning model that is customised to predict the closest target words. It represents words as dense vectors and preserves semantic relationships between words. BERT is a language state-of-the-art model based on a multi-layer coder-transformer. BERT captures deep contextual and semantic information by considering the full context of a word in a sentence, highly effective in capturing both local and global context, making it superior to TF-IDF and Word2Vec in many

NLP tasks. Training this type of model requires either a huge amount of data, time and computational power, or fine-tuning of the pre-trained model. A fine-tuned BERT model was recently proposed based on data from PHA documents for tree text classification tasks: likelihood categorisation, severity level, and potential consequences (Macêdo et al., 2022). In this case, BERT fine-tuning involved adding an input layer of industry data on top of the trained model as a kind of filter. In industrial safety from 2020, an increasing trend can be traced in the use of Deep Learning, NLP and Text Mining for classification of accident reports, entity extraction, injury or accident prediction and semantic search. BERT is most often applied for classification tasks, semantic search, database cleansing and visualisation, while Word2Vec is also applied for topic modelling (Ricketts et al., 2023). The advantage of the TF-IDF (term frequency-inverse document frequency) approach over BERT and Word2Vec is its simplicity, computational efficiency and transparency. TF-IDF is computationally inexpensive and serves as a starting point for similarity detection in data. However, one of the limitations of TF-IDF is that it doesn't consider semantic meaning, in contrast to BERT and Word2Vec (Cahyani et al., 2021). To overcome this limitation co-occurrence network analysis was used. A co-occurrence network is another well-known technique applied to identify statistical correlations between words (or n-grams) in a corpus of unstructured text. The main components of the network are nodes, which represent keywords, and edges, which show the frequency of matches between nodes in a particular unit of text to visualise potential relationships and clusters. In general, the construction of such a network favours its transparency, its ability to roughly preserve contextual information, highlight keywords, giving a thesis view of the text structure (Liu et al., 2021).

2.1 Raw data pre-processing

For the initial data analysis, the transparency of TF-IDF was crucial in choosing between the methods listed above. After initial data cleaning, 865 out of 891 records were selected for work. All data manipulations were performed in Python. The data is anonymous and, by the end of cleaning, contains information only covering age categories of employees, incident time, and type of operation being performed at the time of the incident. The severity of the accident was determined by the number of recovery days (absence at work), which were categorised into intervals of up to one week, two weeks, a month, more than 30 days, or permanent disability (Table 2). The main data for analysis were the 'description' and 'operation' columns, which contained a description of the event that occurred, the injured body part, and in some analyses the cause and type of operation. It should be noted that textual data were translated from Italian to English only at the last steps of the pre-processing phase, to minimise biases and meaning loss of industry-specific terms (e.g. 'rizzaggio' or 'ralla').

Table 2: Initial data after validation and cleaning (first five sample reports).

Id	Age group	Time	Recovery	Operation	Description	Severity
0	4	16.0	7	Movimentazione container	Durante le operazioni...	7-14
1	6	8.0	5	Imbarco merci	Mentre si svolgeva l'operazione...	1-7
2	5	8.0	7	Sbarco container	Nelle operazioni di sbarco…	7-14
3	4	0.0	3	In itinere	Dirigendomi a lavoro, nel percorso...	1-7
4	2	2.0	7	Apertura e rimozione twistlocks	Attuando le operazioni di apertura e…	7-14

Before vectorisation, the text was properly exposed to several pre-processing procedures.

1. Text normalisation by case folding and lemmatization. Case folding aims at reducing the dimensionality of the dictionary and it is not detrimental to distortion of the meaning of the source text. Lemmatization is the normalisation of a word to its semantic root, the lemma, using a knowledge base of synonyms and word endings to combine only words close in meaning into a single token (distinguishing 'developer' from 'development'). Commonly acknowledged libraries for natural language processing are spaCy and NLTK, including a set of rules, a lexical dictionary and statistical models and used for lemmatization.
2. In this study, the 'it_core_news_sm' model adapted from spaCy, was utilised for Italian text processing.
3. Text tokenization, which is most often represented by text segmentation into separate words (in some cases it is more appropriate to break sentences into n-grams). From the NLTK library, the TweetTokenizer tokenizer was used, which was developed to process texts from social networks. In particular, taking into account some abbreviations and emoticons in the text, it allows saving them into a separate token (':))' rather than ':' ')' ')'), which in some cases helps to preserve the emotional colouring of the text content.
4. Filtering tokens by removing stop words, i.e. common language words not carrying a semantic meaning.

2.2 Vectorized text visualisation

The filtered tokens of the 'description' and 'operation' columns were fed into the TfidfVectorizer model and classified according to four recovery categories: 1-7 days for recovery, 7-14, 14-30 and 30+. The first category (1-7) is the most numerous and the last one (30+) is much less represented compared with others (less than 6% of cases). Visualisation of the resulting splitting for the 'description' and 'operation' columns was done by creating Word Clouds shown in Figures 2 and 3 respectively for low and middle severity categories. Word clouds are typically used to visualise words from text based on their frequency occurrence. However, by using TF-IDF vectorisation, each word is assigned a numerical value that reflects its importance within the context of the document and the corpus, as a whole. According to this approach, it is easier to recognise the relationship between the severity of the injury, the work activities, the injured body part and the cause of the injury.

2.3 Co-occurrence network analysis

For the same 'description' and 'operation' columns, combinations of word pairs were created and the 300 most frequent ones were sorted (for visibility purposes). The word pairs were also translated into English with refinement of specific terms such as 'ralla' (part of the attachment of the container to the truck). For convenience, the co-occurrence network was visualised in Gephi software. Each node of the network represents a single word from selected and merged pairs of columns with description and operation. The weight of edges between nodes expresses the number of matches of word pairs. Parameters such as average degree (the average number of edges connected to each node in the graph) and modularity class metrics (assesses the quality of partitioning the graph into communities and indicates how well nodes within one group are connected to each other compared to how they are connected to nodes in other groups) were used to identify and cluster the network nodes. Colour categorisation by modularity helps to reveal the structure of links between words in a text: for example, which words are more likely to occur together and can be related by meaning and context.

3. Results and discussion

It stands to reason that from a resilience perspective, a company's main resource is represented by employees. The severity of injuries and the number of recovery days they entail are considered here as one of the metrics of resilience. Categories '1-7' (most frequent one) and '30+' (highest severity to the system) are the most noteworthy. The results of the importance of particular tokens for each 'description' and related 'operation' for each recovery category are graphically displayed in Figure 2 and 3. Only the first 20 description and operation tokens after TF-IDF were visualised in word clouds to attain a better interpretation of the result.

Figure 2: TF-IDF Word Clouds for 1-7 and 7-14 severity categories.

Figure 3: TF-IDF Word Clouds for 14-30 and 30+ severity categories.

The results are presented in Table 3 in descending order of each word weight for all the selected categories.

Table 3: Word ranking by importance for each severity category.

Severity [days]	Description	Operation
1-7	'fall', 'right', 'contusion', 'left', 'bump', 'hand', 'stumble', 'knee', 'half', 'hit'	'unloading', 'cntrs', 'boarding', 'fastening', 'defastening', 'outside', 'trailers', 'goods', 'handling', 'baggage'
7-14	'fall', 'hand', 'right', 'left', 'fracture', 'half' 'bump', 'stumble', 'foot', 'contusion',	'cntrs', 'unloading', 'boarding', 'fastening', 'outside', 'defastening', 'handling', 'trailers', 'goods', 'half'
14-30	'fall', 'left', 'right', 'hand', 'contusion', 'half', 'stumble', 'fastening', 'bump'	'cntrs', 'unloading', 'loading', 'defastening', 'fastening', 'outside', 'goods', 'trailers', 'handling', 'half'
30+	'right', 'fracture', 'left', 'fall', 'foot', 'stumble', 'fall', 'ground', 'ankle', 'half'	'fastening', 'outside', 'cntrs', 'boarding', 'unloading', 'defastening', 'goods', 'handling', 'stay', 'trailers'

Within the recovery category '1-7', injuries are mainly related to falls, contusions, and right-hand injuries during the activities of container unloading from the vessel, loading onto the ship and fixing the container on the vessel. Severe injuries (more than 30 days) like fractures are mainly connected with fastening containers and transfer. It should be underlined that the here developed methodology is a way to detect weak signals of a resilience system, but not a technique to extract immediate cause or statistical figures regarding accident form and material agent. The next step is to create a co-occurrence network, where we try to find the connected components of the most significant weak signals obtained from tf-idf vectorisation. Figure 4 depicts a possible narrative splitting into the main 5 clusters (by colours), where the node names are proportionally equal to the frequency of occurrence in the selected word pairs. As previously anticipated, colour categorisation highlights the context structure of links between words in a text. Word pairs from the 'description' and 'operation' columns were involved in the network construction, but this time not divided by categories of severity.

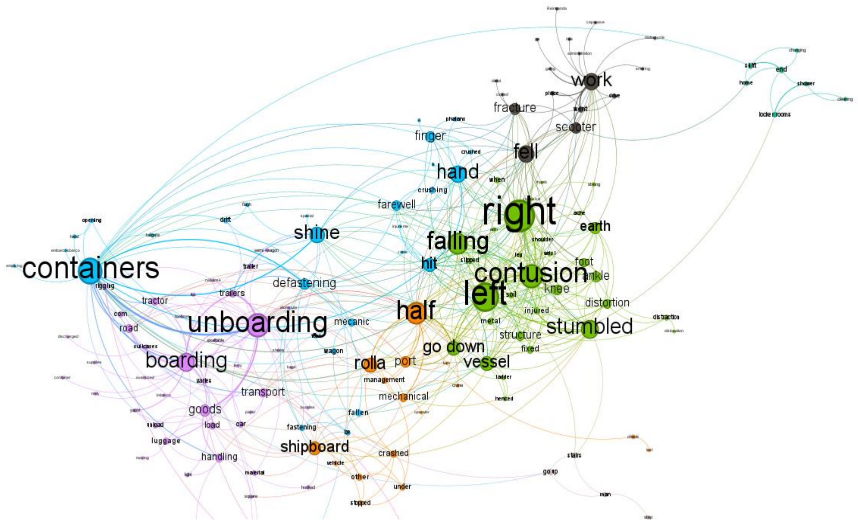

Figure 4: Co-occurrence network for ‘description’ and ‘operation’.

Selecting categories '1-7' and '30+' as the most noteworthy for resilience, the co-occurrence network links of the most significant tokens of ‘description’ and ‘operation’ from TF-IDF word clouds was designed.

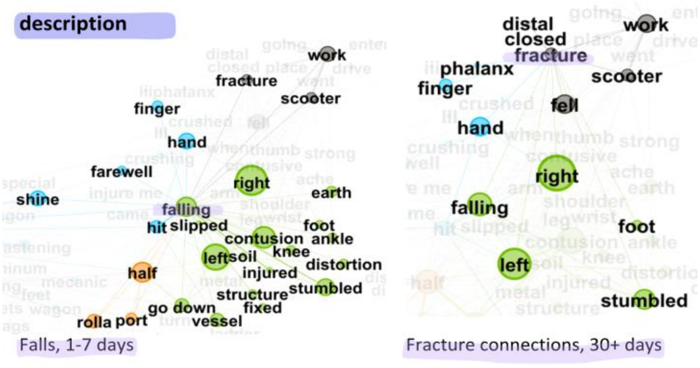

Figure 5: Highlighting relevant network components for the item 'description'.

Figures 5 and 6 show related items for falls and fractures from ‘description’, unloading and fastening from ‘operation’ respectively. Considering the '1-7' ranking, fall tokens are associated with container unloading operations, to a minor extent with unloading goods. The right hands are the most often injured part in these operations. For '30+', fractures are related to right hand injuries in road traffic accidents on the way from work. Additionally, incidents are associated with the fastening of trailers. Upon proper refinement, it will be possible to implement into the resilience system the learning ability to retrieve from the unstructured text of accident reports the unclassified pre-categorised accident cause and form, useful as warnings of future issues, including emerging hazards in the port environment (Pasman et al., 2023). The information, possibly connected with on-field retrieved data showing interdependencies between the components of the infrastructure, can support the resilience system flexibility as a function of anticipation in the reality of man-new-hazard interaction.

Figure 6: Highlighting relevant network components for the item 'operation'

5. Conclusions

This study outlines a designed analysis framework of unstructured data from accident reports useful to identify notable operations requiring special attention and support, as well as possible shortcomings in safety management systems. TF-IDF is a trustworthy, not black box tool to prioritise the importance of words in reports, while the co-occurrence network captures relationships between words based on their proximity in the text, which adds contextualisation to the analysis. These TM methods can help the resilience system to highlight its vulnerabilities in dynamic progression, especially with the active involvement of employees, who can provide the system with more data on current situations and emerging hazards. The further step is fine-tuning of BERT model on broader data sets (e.g. shift reports, surveys on work environment actual conditions, equipment failure and near-miss reports) for identifying new optimization opportunities and improving learning and predictive capacity of the resilience system by semantic information capable of predicting the complex system behaviour.

Acknowledgments

Funded by the European Union - NextGenerationEU and by the Ministry of University and Research (MUR), National Recovery and Resilience Plan (NRRP), Mission 4, Component 2, Investment 1.5, project “RAISE - Robotics and AI for Socio-economic Empowerment” (ECS00000035).

References

Bragatto P., Vairo T., Milazzo M. F., Fabiano B., 2021, The impact of the COVID-19 pandemic on the safety management in Italian Seveso industries, Journal of Loss Prevention in the Process Industries, 70, 104393.

Cahyani D.E., Patasik I., 2021, Performance comparison of TF-IDF and Word2Vec models for emotion text classification, Bulletin of Electrical Engineering and Informatics, 10, 2780 – 2788.

Dinh L. T. T., Pasman H., Gao X., and Mannan M. S., 2012, Resilience engineering of industrial processes: principles and contributing factors, Journal of Loss Prevention in the Process Industries, 25, 233-241.

Fabiano B., Guastaferro M., Pettinato M., Pasman H.J., 2024, Towards strengthening resilience of organizations by risk management tools: A scientometric perspective on COVID-19 experience in a healthcare and industrial setting, The Canadian Journal of Chemical Engineering, 102, 5, 1705-1725.

Fraccascia L., Giannoccaro I., Albino V., 2018, Resilience of complex systems: state of the art and directions for future research, Complexity, article ID 3421529 DOI:10.1155/2018/3421529.

Liu G., Boyd M., Yu M., Quddus N., 2021, Identifying causality and contributory factors of pipeline incidents by employing language processing and text mining techniques, Process Saf. Environ. Protection, 152, 37-46.

Macêdo J. B., Moura M., Aichele D., Lins I.D., 2022, Identification of risk features using text mining and BERT-based models: Application to an oil refinery, Process Safety and Environmental Protection, 158, 382-399.

Pasman H.J., Sripaul E., Khan F., Fabiano B., 2023, Energy transition technology comes with new process safety challenges and risks, Process Safety and Environmental Protection, 177, 765-794.

Ricketts J., Barry D., Guo W., Pelham J., 2023, A scoping literature review of Natural Language Processing. Application to safety occurrence reports, Safety, 9, 22.

Vairo T., Agnello P., Currò F., Fabiano B., 2023, Unveiling the Achilles’ heel: detecting organizational weaknesses in the energetic transition challenge, Chemical Engineering Transactions, 105, 499-504.

Vairo T., Gualeni P., Reverberi A.P., Fabiano B., 2021, Resilience dynamic assessment based on precursor events: Application to ship lng bunkering operations, Sustainability, 13 (12), 6836.