Data-driven Mapping Between Proactive and Reactive Measures of Occupational Safety Performance

Chapter · January 2018

DOI: 10.1007/978-981-10-6328-2_5

4 authors:

Abhishek Verma
Indian Institute of Technology Kharagpur

Subit Chatterjee
Indian Institute of Technology Kharagpur

Sobhan Sarkar
Indian Institute of Technology Kharagpur

Jhareswar Maiti
Indian Institute of Technology Kharagpur

Chapter 5
Data-driven Mapping Between Proactive and Reactive Measures of Occupational Safety Performance

Abhishek Verma, Subit Chatterjee, Sobhan Sarkar and J. Maiti

Abstract This study aims to analyse the incident investigation reports logged after the occurrence of events from an integrated steel plant and map it with proactive safety data. From the narrative text describing the event, this study has attempted to unfold the hazards and safety factors present at the workplace. Text document clustering with expectation maximization algorithm (EM) has been used to group the different events and find key phrases from them. These key phrases are considered as the root causes of the reported events. This study shows how the mapping of the safety factors from both proactive safety data and incident reports can help in the improvement of safety performance as well as better allocation of resources. The study points out specific areas to the management where improvements are needed. The mapping also indicates the areas of improvement made by the constant effort of safety practitioners.

Keywords Incident reports · Proactive safety data · Text document clustering

5.1 Introduction

Steel is essential for the development of any economy because its usage ranges from household products to complex industrial and defence machinery. Major industrial economies are defined and strengthen the growth of strong steel industries. Indian economy is also dependent on the growth of steel industries.

A. Verma (✉) · S. Chatterjee · S. Sarkar · J. Maiti
Department of Industrial & Systems Engineering, IIT Kharagpur, Kharagpur, India
e-mail: abhishekverma.cs@gmail.com

S. Chatterjee
e-mail: subitme16@gmail.com

S. Sarkar
e-mail: sobhan.sarkar@gmail.com

J. Maiti
e-mail: jhareswar.maiti@gmail.com

J. Maiti and P.K. Ray (eds.), *Industrial Safety Management*,
Managing the Asian Century, https://doi.org/10.1007/978-981-10-6328-2_5

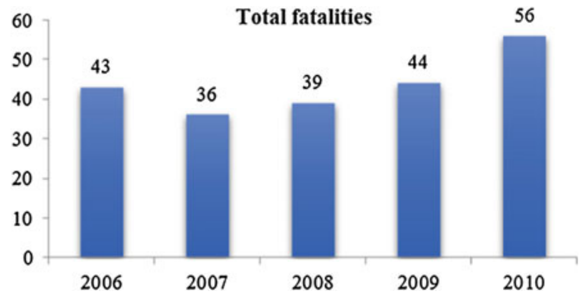


Fig. 5.1 Year-wise distribution of fatalities occurred in major steel producers of India

India holds a key position in world steel production map. Steel sector contributes about 2% to Indian GDP. As per the global scenario, India is the fourth largest producer of crude steel with about 88mt production in the year 2014–2015 (Ministry of Steel, India). Since the production demand is increasing and there is a requirement of building more infrastructure (e.g. workforce, technology, machinery, roads) with the constraints (e.g. space, environmental law, and regulation), which are making the workplace hazardous. The whole process of steel production from raw material to finished products exposes the workers to the wide range of physical (noise, heat, vibration, slip/trip/fall, etc.), chemical (gasses, fuses, etc.) and biological hazards (Jovanovic et al. 2004). In the report of the working group on steel, India for the 12th five-year plan published the year-wise fatalities occurred in major steel producing industries shown in Fig. 5.1. The figure indicates that despite considerable advancement fatalities are increasing over the years. So, ensuring safety at the workplace with keeping the pace of production rate is the topmost priorities for all the industries.

Most of the research efforts in safety management have been focused primarily on analysing and investigating the past accidents but now attention is being directed towards proactive measures to protect the employees and enrich safety culture (Sheehan et al. 2016). Routine safety observation activity can be considered as lead indicators (Dyreborg 2009). In managerial arena there is an increased demand and interest in encouraging firms to use proactive signals rather than relying on lag indicators (incidents/accidents) (Sinelnikov et al. 2015). The importance of proactive safety measures and their relationship to overall business performance are well understood and accepted by senior management. This relationship is pivotal for a proactive strategy in safety management. For this purpose, many of industries are collecting proactive and reactive data.

In the last decade, the data collection in the safety field increases after the introduction of online safety management system database. Root cause extraction from this vast data remains at the core of Safety Management System (SMS) to gather useful information to target the particular system fault because information is captured in both structured and narrative text format. The structured part of data tells about the scenario up to some convincing extent but gets completed by including the narrative text data, at least presumptively. Most of the organizations collect data about the unsafe act, unsafe conditions, and other hazards (lead indicators) in the form of free text. Free text description provides the detail about the hazard, such as the description of the machine, exact location, surrounding condition, to describe the situation prevailing in the plant. Similarly, whenever an incident happens, a brief description of the incident (lag indicator) is reported as incident reports in safety management system (SMS) of the organization in the form of free text. But it also increases the complexity for an analyst to extract the information from that because free text provides the freedom to explain the incident in their words which results in noisy text generation.

To extract hidden information from the vast amount of unstructured data, higher management is searching for methods for taking a smarter decision to improve the safety performance. So, efficient techniques are needed to identify the cause to improve the safety performance in all the safety critical industries. These hidden 'knowledge nuggets' from the large volume of data are practically undetectable using traditional tools and techniques and can be discovered by using advanced techniques of data mining and machine learning (Watson 2008) .

The text mining can support the extraction of lethal factors and pattern from a large volume of data, undiscoverable with the traditional methods. Text clustering analysis is a method for exploration and visualization of textual data, to fully understand the information and the structure of original text documents. Text document clustering is an essential text mining utility which enables the discloser of recurring events. Text clustering analyses have preliminarily shown its performance and usefulness as compared with traditional tools and techniques (Saraçoğlu et al. 2008). For this purpose researchers started using the software with advanced tool and techniques combined with expert analysts (Cleary 2011).

This study introduces safety data analysis as an application area. Steel plant considered for this study captures the hazards and incident related information using an IT enabled safety management system. In this research text document clustering was utilized to study the incident investigation reports and safety observation data to find out root causes and trend, without any prior assumption of their presence. The main target behind the use of clustering techniques is to explore and identify the weak signals from the free text description of incidents and hazards.

Expectation maximization clustering algorithms were used to group loosely related reports and documents, which explains the process of incidents combining the human, technical and organizational factors from incident reports and hazards events from proactive data. It also identifies interconnected reports and discovers the possible recurring irregularities, to determine the hot spots of high risks.

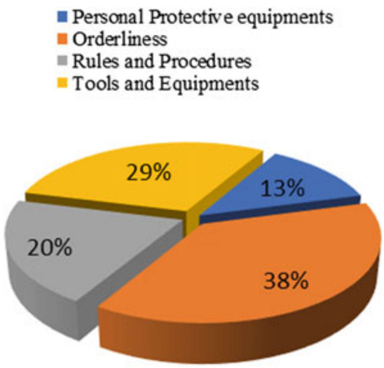
In this study, we propose a methodology for mapping the proactive and reactive safety measure to improve the overall safety performance of the plant. The rest of the paper is organized as follows. Section 5.2 describes data collection preparation. Section 5.3 presents the conceptual research methodology adopted for this study. Results and discussion are given in Sect. 5.4. Conclusions and limitation of the study are provided in Sect. 5.5.

5.2 Data Collection and Preparation

The current study was conducted using incident investigation (reactive) and safety observation data (proactive) of the iron making (IM) division of an integrated steel plant of India. Any employee involved in incident or witness of any hazardous situation at workplace can report the same to the corresponding supervisor of the department or log in the SMS. Whenever worker notices any unsafe act or unsafe condition that has the potential of causing an accident, these observations are recorded into the SMS of the organization. The safety observation data collected under four categories: (i) improper tools and equipment, (ii) rules and procedures, (iii) personal protective equipment (PPE), and (iv) orderliness. The brief description of the hazardous behaviour or condition is noted as free text to narrate the complete scenario. This free text in the form of 'brief description' is an unexplored area and our primary focus for our study. Similarly, whenever incidents happen, employee logs every detail of the incident in the SMS. The incidents are mainly reported under three incident categories: (i) injury/property damage, (ii) medical cases, and (iii) near-miss incidents. These incidents may end up in different impacts. Fifteen different impacts for the various incidents were listed in the incident reports which are: (i) equipment property damage, (ii) derailment, (iii) first aid, (iv) fire, (v) LTI, (vi) exgratia, (vii) toxic Release, (viii) uncontrolled environment, (ix) fatality, (x) medical ailment (major), (xi) medical ailment (minor), (xii) foreign body, (xiii) radio activity, (xiv) death, and (xv) injury on duty. The complete process of incident is narrated in free text form to that can help practitioners in getting insights that are primarily responsible.

Both the safety data for IM division were collected for 17 months (April 2014–August 2015). Figures 5.2 and 5.3 shows the distribution of safety observation and incident data respectively. Figure 5.2 shows that most the observations were made for issues related to orderliness (38%) and tools and equipment (29%).

Fig. 5.2 Safety observation data distribution



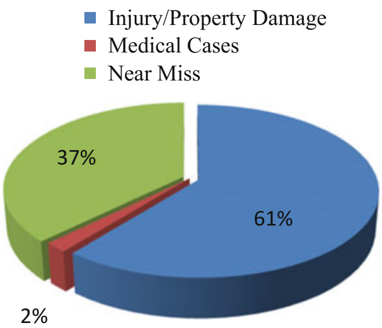


Fig. 5.3 Incident data distribution

Figure 5.3 shows that accidents (61%) are logged more than near-miss (37%) cases. This indicated that worker might not be aware of importance of near-miss reporting.

This study focuses on analysing the descriptive text data. Unfortunately, this section of the report contains lots of inconsistency. So, for the success of any data mining or text mining algorithm, data preparation stage is a deciding factor to extract the quality information (Freitas 2002). Almost 80% time is utilized in pre-processing and preparing the data in any data mining or machine learning project (Zhang et al. 2003). In text mining, data pre-processing is required to remove noise and irrelevant information for improvement in the quality of data (Rajman and Vesely 2004). Data preparation involves various issues related to text data such as a spelling error, non-vocabulary words (Hindi words), incomplete information, irrelevant information (names and address). Duplicate data rows were removed using remove duplicate tool of MS Excel itself. Misspelling and irrelevant shorten text was removed using MS Excel function and manual review.

5.3 Methodology

After data preparation, the variables of interest are extracted from incident reports for further analysis. In our study, the variables of interest are 'incident category' & 'brief description of incident' from incident data and 'observation category' & 'brief description of observation' from safety observation data. In the narrative text data analysis stage, text document clustering technique is utilized to extract the hidden factors in the form of descriptive terms. The descriptive terms from observation data can act like proactive (lead) indicators and descriptive terms form incident data will act as root causes or lag indicators. Then we will try to map all the lagging indicator to corresponding lead indicator. The link will indicate that the lag indicator was already observed as a hazard during safety observation visits. Unlinked lag indicator indicates that there are almost no hazardous observations (lead indicators), logged in past that can be treated as the causal factors.

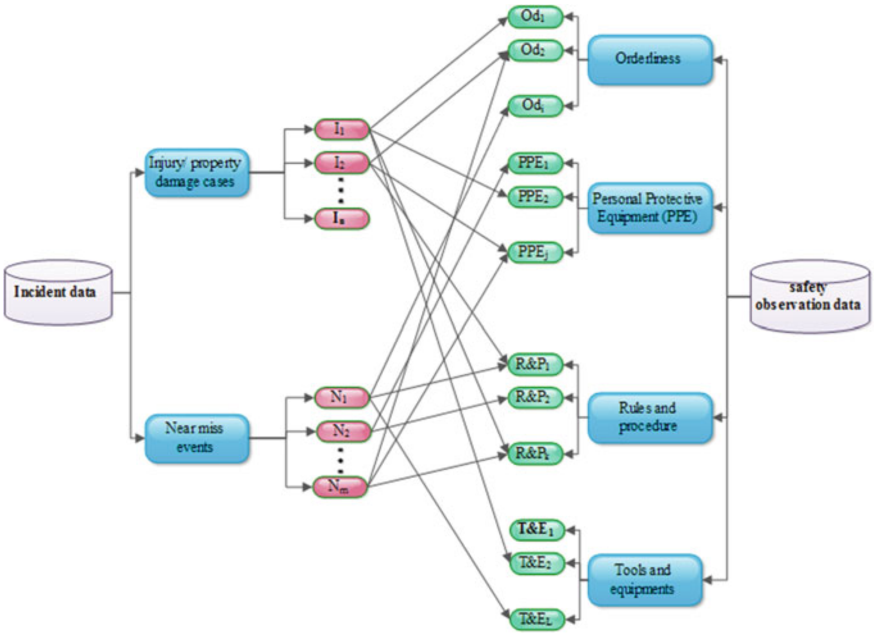


Fig. 5.4 Visualization of mapping the root causes of accident data with safety observation data

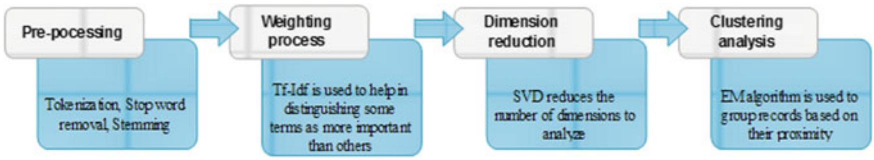


Fig. 5.5 Schematic diagram for text document clustering

Unlinked lead indicator suggests that those hazardous events were successfully mitigated and not converted into any kind of consequences. A visualization of mapping the root causes of incident data with safety observation data has been shown in Fig. 5.4.

The current study utilized SAS text miner software as a text mining tool. The steps performed for text document clustering are shown in Fig. 5.5.

5.4 Results and Discussion

The incident data collected from the SMS had injury/property damage, medical cases and near-miss incidents. Medical cases consist of only 2% of the total incidents. So, medical cases were combined with injury/property damage data for text clustering. Text clustering was performed for accidents (injury/property damage, medical cases) and near-miss incidents separately. With the help of expert advice having wide knowledge of safety domain in steel plant, the root causes/keywords have been inferred from the clustering result. Keywords from the incident clustering result will explore the root causes of accidents. Clustering of injury/medical cases gave 14 clusters and clustering of near-miss data gave eight clusters. Some of the clusters have been combined due to redundancy. Figures 5.6 and 5.7 show the 14 clusters for 'injury/property damage + medical cases' and seven clusters for near-miss incident cases since one cluster of near-miss is redundant. Due to lack of space, the detailed cluster is not provided in the paper (Table 5.1).

It can be seen from the clustering output of injury/property damage + medical cases that wagon derailment cases appear in many cluster (e.g. sinter plant wagon derailment, wagon derailment, quenching loco derailment, and loco torpedo derailment). So these clusters have been clubbed into one broad cluster of 'derailment'. Similarly, 'motorcycle accident' and 'vehicle skidding' clusters are

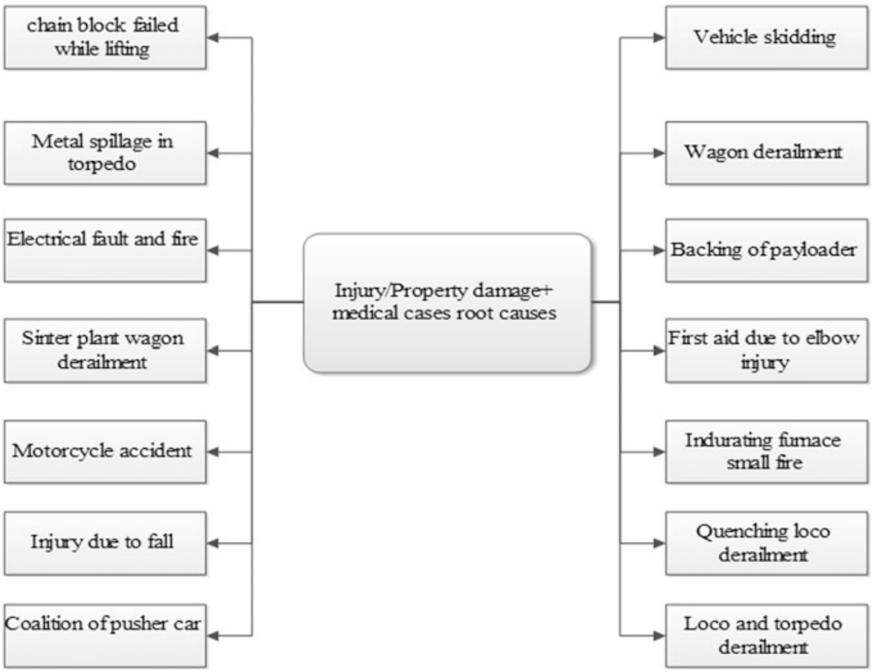


Fig. 5.6 Root causes of injury/property damage + medical cases

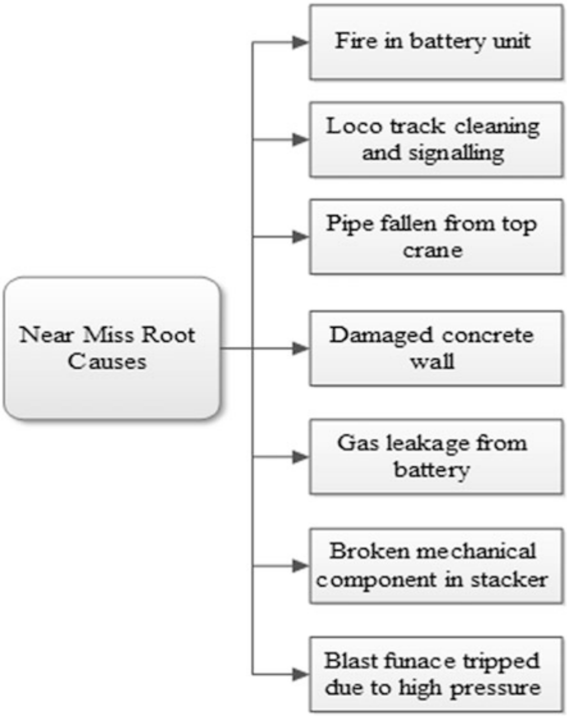


Fig. 5.7 Root causes of near-miss incident

Table 5.1 Number of Text document clusters found for different categories

Sl. No.	Type of data set	Categories	Number of clusters
1	Reactive	Injury/property damage + medical case	14
2		Near-miss	7
3	Proactive	Orderliness	22
4		Personal protective equipment	6
5		Rules and procedure	12
6		Tools and equipment	16

closely related and has been clubbed into one general cluster of ‘motorcycle accident’.

Similarly, the safety observation data collected from the SMS had subcategories such as orderliness, personal protective equipment, rules and procedures and tools and equipment. Text clustering was performed on each of these different subcategories separately with the same system settings as used in the incident clustering. Keywords from the clustering result will give us insight into the hazardous situations noted by safety officials during their site visit. However, it was not always possible to decipher keywords from all the clusters. So those clusters have been neglected.

5.4.1 Mapping of Incident Root Causes Versus Safety Observations

To identify whether the underlying root causes bringing information about the incidents had been identified as potential risks during safety observation of workplaces, detailed mapping of the root causes of incidents against the lead indicators of hazards noted during safety observation has been done as shown in Fig. 5.8 for injury/property damage + medical cases and in Fig. 5.9 for near-miss incidents.

Few of the underlying root causes such as electrical fault, injury due to fall, coalition of the pusher car in the battery unit, wagon derailment have been looked upon during site visits as potential risks. On the other hand, few of the root causes of accidents such as 'metal spillage in torpedo', 'backing of pay loader', and 'chain block problem' were not anticipated as potential hazards during safety visits. The safety authorities overlooked or neglected to see these as potentially dangerous circumstance. Legitimate moves ought to be made so that these underlying root causes are not disregarded any further and preventive moves ought to be made to moderate any conditions emerging from these root causes. Wagon derailment was found as the most basic cause of accidents. But only a few safety observation data were reflecting the root cause of wagon derailment. Few incidents were reported for the failure of chain pulley or frame failure. However, no safety observations were

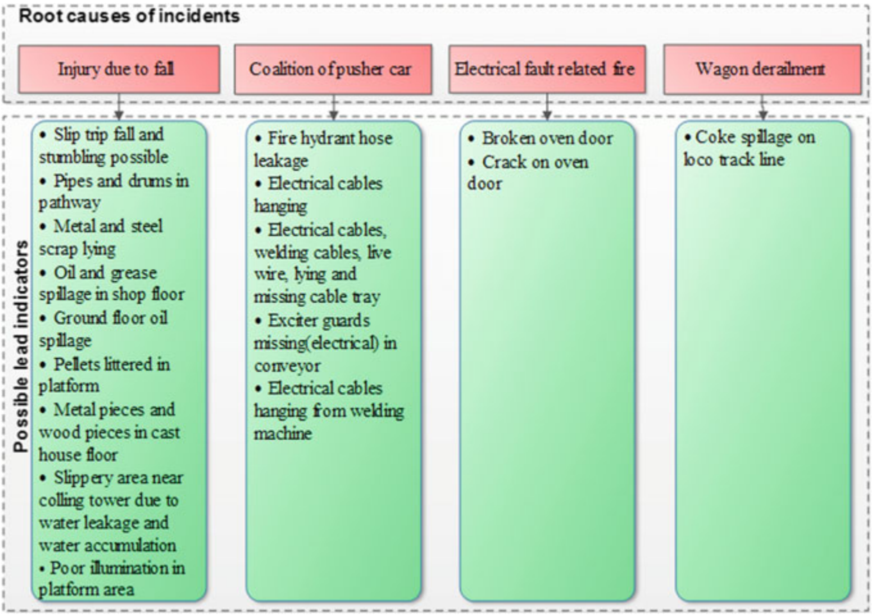


Fig. 5.8 Mapping of root causes of injury/property damage + medical cases with the safety observation data

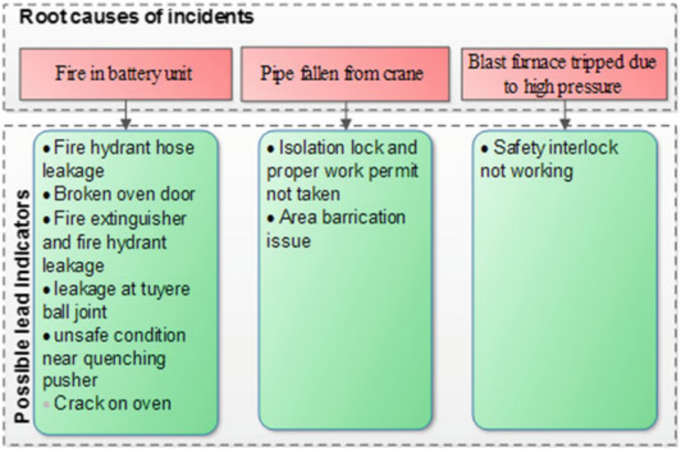


Fig. 5.9 Mapping of root causes of near-miss incidents with safety observation data

logged regarding the checking of chain pulley or other lifting equipment. Third party certification is necessary for lifting equipment. However, none of the safety observations took note of the availability of these certifications of lifting equipment before they were put to work in site.

For the near-miss incidents, many incidents were found due to some issues in the battery unit. The safety observation data show that officials take care of hazardous condition in the battery unit. Leakage of gases due to mechanical failure, broken oven door, the problem with the pusher door was taken care of during the site visit.

Preventive measures were taken to mitigate any fire incident happening in the plant. Fire extinguisher, fire hydrant hose leakages were checked in case of any fire related incidents. Similarly, to avoid falling related incidents, barrication, PPE's of staff working, proper work permit was checked. But the root causes such as loco signalling and damaged concrete wall could not be mapped with and any lead indicators.

5.5 Conclusion

This study tried to consider the mapping between root causes of incidents with lead indicators extracted from the narrative text data. The root causes show the underlying factors behind various incidents and lead indicators shows the focusing area where safety officials are searching for possible hazards. Text document clustering used for analysis of incident and safety observation data provided us with insights into the major accidents happening in the steel plant and their root causes. However, it is not sufficient to showcase all the event description, but it helped in pointing out the areas where the organization needs to improve their safety framework. We have outlined several lead indicators for measurement and monitoring of industrial safety. We are hopeful to conclude here that it can provide practitioners with useful information to support the organization to move away from a focus on lag indicators towards a preventative focus on lead indicators.

Acknowledgements The authors gratefully acknowledge the management of steel plant for their support and cooperation.

References

Cleary, D.: Predictive Analytics in the Public Sector: Using Data Mining to Assist Better Target Selection for Audit. Electron. J. e-Government **9**(2), 132–140 (2011)

Dyreborg, J.: The causal relation between lead and lag indicators. Saf. Sci. **47**(4), 474–475 (2009)

Freitas, A.A.: Data mining and knowledge discovery with evolutionary algorithms, 1st ed. Springer, Heidelberg (2002)

Jovanovic, J., Arandelovic, M., Jovanovic, M.: Multidisciplinary aspects of occupational accidents and injuries. Work. Living Environ. Prot. **2**(4), 325–333 (2004)

Rajman, M., Vesely, M.: From text to knowledge : document processing and visualization : a text mining approach. Text Mining and Its Applications: Results of the NEMIS Launch Conference. Springer Science & Business Media, pp. 7–24 (2004)

Saraçoğlu, R., Tütüncü, K., Allahverdi, N.: A new approach on search for similar documents with multiple categories using fuzzy clustering. Expert Syst. Appl. **34**, 2545–2554 (2008)

Sheehan, C., Donohue, R., Shea, T., Cooper, B., De Cieri, H.: Leading and lagging indicators of occupational health and safety: the moderating role of safety leadership. Accid. Anal. Prev. **92**, 130–138 (2016)

Sinelnikov, S., Inouye, J., Kerper, S.: Using leading indicators to measure occupational health and safety performance. Saf. Sci. **72**, 240–248 (2015)

Watson, R.T.: Data Management, Databases and Organizations. Wiley, New York (2008)

Zhang, S., Zhang, C., Yang, Q.: Data preparation for data mining. Appl. Artif. Intell. **17** March 2015, 375–381 (2003)