Public Awareness of Traffic Safety based on Text Analytics

S. S. Syed Ahmad[*1], M. Muhammad[2], Z. M. Jawi[3], E. Kasmuri[1] and N. Z. Zulkarnain[1]

[1]Center for Adv. Computing Technology (C-ACT), Fac. of Information & Comm. Tech., Universiti Teknikal Malaysia (UTeM), 76100 Durian Tunggal, Melaka, Malaysia
[2]Aibots Sdn. Bhd., Menara Zurich, Taman Abad, 80300 Johor Bahru, Johor, Malaysia
[3]Malaysian Institute of Road Safety Research, 43000 Kajang, Malaysia

[*]Corresponding author: sakinah@utem.edu.my

ORIGINAL ARTICLE **Open Access**

Abstract – *Traffic and road safety is a major concern of many governments around the world as it ensures the safety of its road users. Due to an increase in the number of road users, traffic-related accidents and fatalities have increased annually. Therefore, governments have stepped up their efforts to ensure the safety of road users and to reduce the number of traffic-related accidents and fatalities. However, these efforts have proven ineffective as the data gathered to measure the effectiveness of these efforts either lack credibility or are insufficient. This has led to inconclusive findings concerning the culture of road users and the extent of their awareness of traffic safety. Therefore, this paper performs a systematic data- and text-driven analytics and review of published literature on public awareness of traffic and road safety. The selected papers were first subjected to content analysis before being grouped according to the frequency of use of keywords discussing the issue at hand. As this method enabled us to successfully identify trends as well as the efficacy of traffic and road safety research, therefore, data and text analytics is an effective method of conducting a systematic literature review. The findings of this study will help traffic authorities gain credible insights into public perception, behavior, and attitude towards traffic safety awareness. Our assessment indicates that there are two important perspectives to be considered with regard to road safety; namely human and technology; the latter of which will be the main topic of study on public awareness of road and traffic safety. Moreover, an emerging trend in the literature suggests that more efforts should be allocated to promoting public awareness of road and traffic safety.*

Keywords: Traffic safety, road safety, text analytics, systematic literature review

1.0 INTRODUCTION

Road safety is regarded by the World Health Organization as one of the fastest-growing threats to human lives across the globe (WHO, 2015). Traffic-related injuries and fatalities result in massive loss of human lives and properties annually. The "UN Decade of Action for Road Safety (2011-2020)", which was adopted in March 2010 stated that approximately 1.35 million lives are lost to traffic-related accidents and that more than 50 million are injured every year (WHO, 2011). The "Make Roads Safe" report by the Commission for Global Road Safety confirms that traffic-related accidents are the primary cause of youth mortality worldwide (FIA Foundation, 2006). The Chair of the Commission for Global Road Safety emphasizes that "one of the reasons the problem has not been addressed is the absence of accountability for road safety at international and domestic levels" (FIA Foundation, 2006). Studies reveal that most traffic crashes are caused by drive-associated factors. This is because road stakeholders sometimes neglect the importance of traffic safety which results in hazardous crashes that result in devastation to themselves and others (UNECE, 2019).

In 2016, Malaysia had the third-highest road accident death rate worldwide, surpassing the world's two most populous countries, China and India (Lum, 2019). The Malaysian Ministry of Transport reported that this trend remained the same between 2009 to 2018 (MOT, 2019). Research conducted by the Malaysian Institute of Road Safety Research (MIROS) concluded that 80% of accidents are caused either by reckless driving or ignoring traffic rules instead of road or vehicle conditions (Wah, 2015). The WHO estimates that among all the countries with vehicle ownership rates above 0.5, Malaysia has the highest fatality rate per 100,000 populations (Figure 1). Although Malaysia has a world-class highway system, which is on par with developed countries, Malaysians have a unique driving culture and mindset when they are on the road, especially during traffic congestion. This includes frustration, road rage, drowsiness while driving, selfishly hogging a lane, incessant weaving, and dangerously forcing their way between other vehicles as well as switching lanes at the last minute to jump the queue. Therefore, traffic safety in Malaysia remains a worrying concern. Though Malaysia has a relatively high proportion of motorcycles on the road, it pales in comparison to the number of motorcycles in neighboring countries such as Vietnam, Cambodia, Laos, Thailand, and Indonesia, where motorcycles account for more than 70% of the total traffic. A possible hypothesis for this situation could be that community learning of road safety may not be as rapid as it is in developed countries despite vast improvements in motor vehicle safety and road infrastructure. This, in turn, may encourage risk compensation, where improved vehicle safety standards and higher geometric design of roads increase the reckless behavior of drivers (Darma, 2017).

In 2016, the UN commissioned a group of high-level consultants to deliberate on sustainable transport. Their published report – Mobilizing Sustainable Transport for Development – contains proposals to promote sustainable transport (UN, 2016). Following these deliberations, the Sustainable Development Goals (SDGs) that were adopted in 2015 highlight the importance of sustainable transport in tackling the social exclusion of vulnerable groups; specifically, Goal 3 – "Good health and wellbeing for people" and Goal 11 – "Sustainable cities and communities", which include goals related to road safety and sustainable transport.
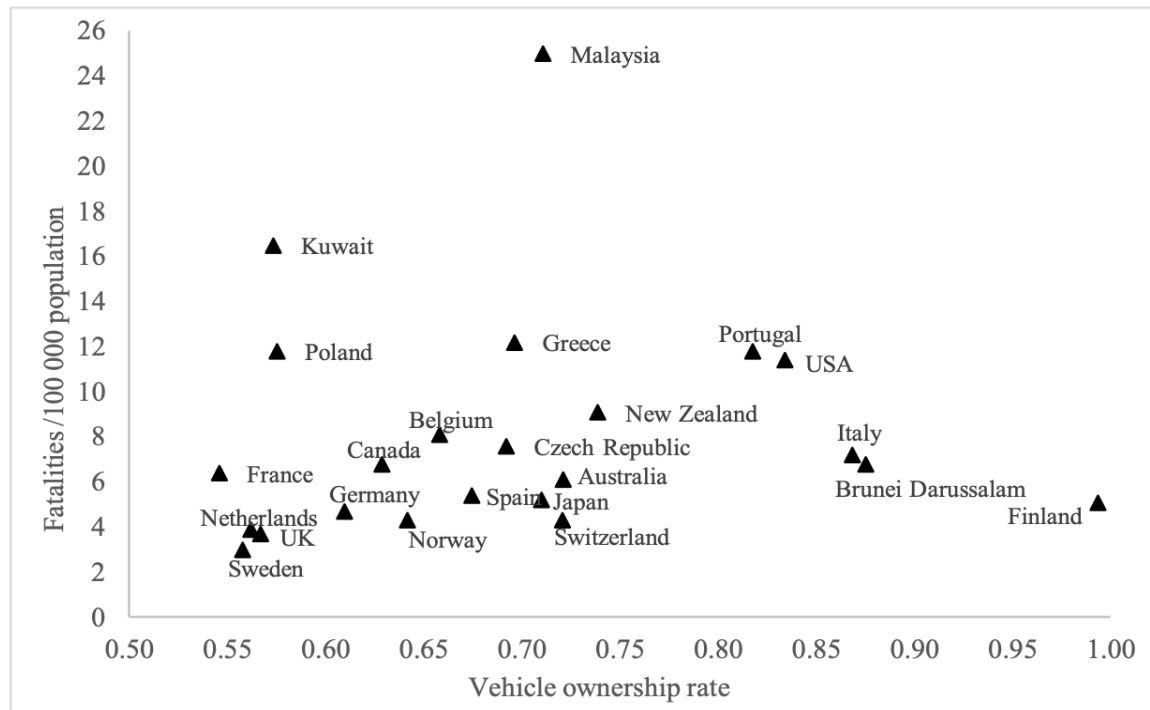
Figure 1: Fatality rate in countries with vehicle ownership rates above 0.5 in 2010
(WHO, 2013)

According to the WHO, 74% of global road traffic accidents occur in low-income countries. When accident statistics in both low and middle-income countries are combined, it accounts for 90% of the global figure. Although most of the global population lives in low and middle-income countries (WHO, 2018), the number of vehicles registered in these countries accounts for approximately 54% of the global total, and yet the number of traffic accidents is disproportionate to the number of vehicles. The fact that accident-related fatalities in high-income countries have decreased between 2000 and 2015 but increased in low-income countries indicates the seriousness of traffic problems in developing countries.

Road traffic injuries cause considerable economic losses to individuals, their families, and nations as a whole. These economic losses arising from the cost of treatment as well as the loss of productivity for those killed or disabled by their injuries. Family members are likewise afflicted if they are required to take time off work or school to care for the injured. Road traffic crashes cost most countries roughly 3% of their gross domestic product (GDP). Figure 2 shows that Brunei and Singapore have low numbers of road deaths which are similar to that of western countries; such as the Netherlands where, although an exception, the number of road deaths is lowest in the world. However, as the number of road deaths in Thailand and Malaysia; both middle-income countries; is extremely high, it suggests that traffic problems have worsened alongside economic growth. Figure 2 also indicates a very high number of road deaths for low-income countries such as Cambodia, Laos, and Myanmar. This supports the "3-5-2" concept; adapted from strategic formations in soccer; that explains the status quo of ten ASEAN countries in certain aspects towards reality-based road safety countermeasures (Jawi & Kassim, 2013). Many factors underlie the rapid increase of traffic accidents, which has become a major social problem. In ASEAN cities, for instance, pedestrians, motorcycles, automobile vehicles, and vehicles unique and native to these countries (e.g., motorcycle taxis and three-wheeler tuk-

tuks) share the road in a chaotic manner. Add to this mix underdeveloped road traffic infrastructure as well as poorly established social institutions; such as the driving license system. Furthermore, measures to educate and improve road safety awareness are neither fully developed nor implemented. These factors intertwine in a complex manner to not only worsen traffic congestion but increase the frequency of accidents. In many ASEAN countries, priority is given to motorcycles and automobile vehicles while pedestrians are often overlooked (Manan, 2014; Hue et al., 2015; Kitamura et al., 2018).
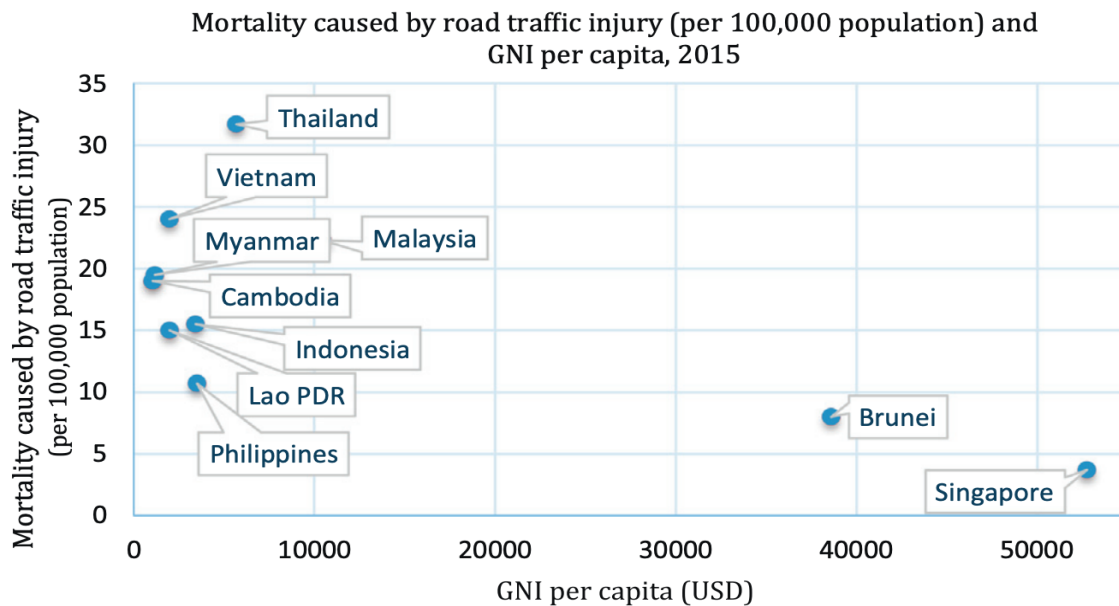

Figure 2: Economic level and traffic-related fatalities in Southeast Asia (2015).
Source: Data from World Development Indicators (WDI)

This study systematically reviews the literature on public awareness of road and traffic safety. To date, very little attention has been given to this topic, particularly among the academic community. Our investigation of the various issues affecting public awareness of road and traffic safety is outlined in this paper. The structure of this paper is as follows: (1) background on traffic safety, (2) an overview of our research methodology, (3) descriptive and text analyses of published research articles, and (5) a summary of the contributions of this study as well as recommendations for future works.

2.0 METHODOLOGY

This study is regarding recent publications on public awareness of traffic and/or road safety. A systematic literature review was conducted to confirm trends as well as investigate the latest research that discusses these issues in depth.

Suitable keywords and search terms relating to the scope of the study were first identified. In an iterative process, the data was gathered over several steps as illustrated in Figure 3. Research published over the past 10 years – between January 2010 to December 2020 – was used in this study as it answers our first objective which is to look at the trends of public awareness of road and traffic safety. Only articles published in English were used in this review as most of the literature is in English. The main keywords used were road safety, traffic safety, and awareness. Boolean operators such as "AND" and "OR" were then used between the

keywords in the search field. For example, in the title, abstract, and keywords: we used ("road safety" OR "traffic safety") AND "public awareness". Table 1 shows the review protocol utilized to provide concrete results on the subject at hand.

Table 1: Review protocol

Item	Description
Keywords	Road Safety; Traffic Safety; Public Awareness
Boolean Operators	(road safety OR traffic safety) AND (awareness)
Search Fields	Title; Abstract; Keywords
Language	English
Publication Type	Article
Time Range (Years)	2010 to 2020
Databases	Scopus (252) and IEEE (64)

A total of 316 papers were obtained after submitting search strings in multiple databases. Text analysis enabled us to select papers that best met the objectives of our research. Of this number, 234 papers remained after duplicate papers were removed. The selected papers were grouped into three categories based on the year of publication. Figure 4 illustrates the papers selected according to the year. This was vital as it showed the trend of research in this field. These publications were examined using text analytics.

Stage 1: Define

Step 1: Identification of need for a literature review
Step 2: Development of a literature review protocol

Stage 2: Collect and Select

Step 1: Identification of documents
Step 2: Selection of relevant documents

Stage 3: Analysis and Findings

Step 1: Categorisation of documents
Step 2: Data extraction
Step 3: Descriptive analysis
Step 4: Textual analysis

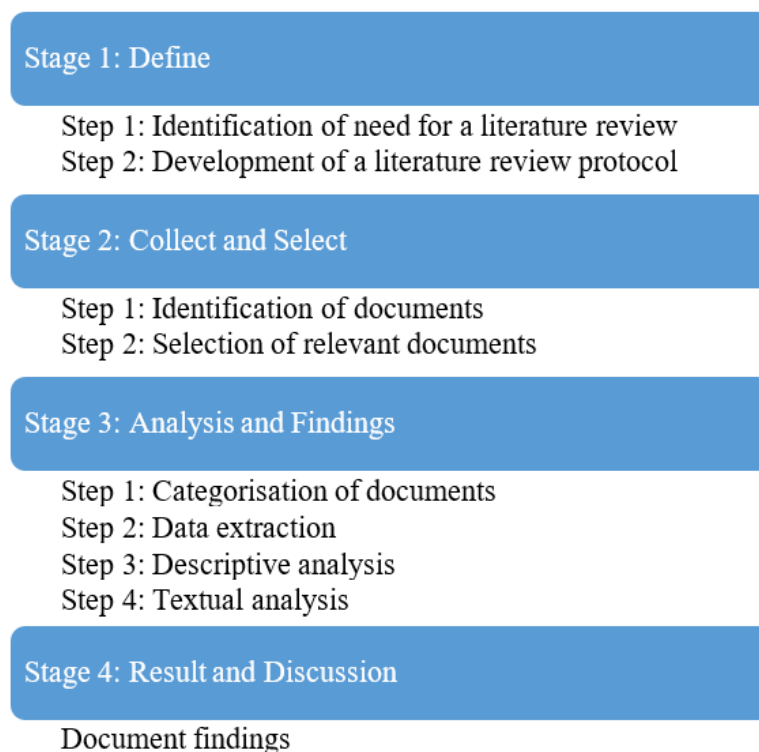**Stage 4: Result and Discussion**

Document findings

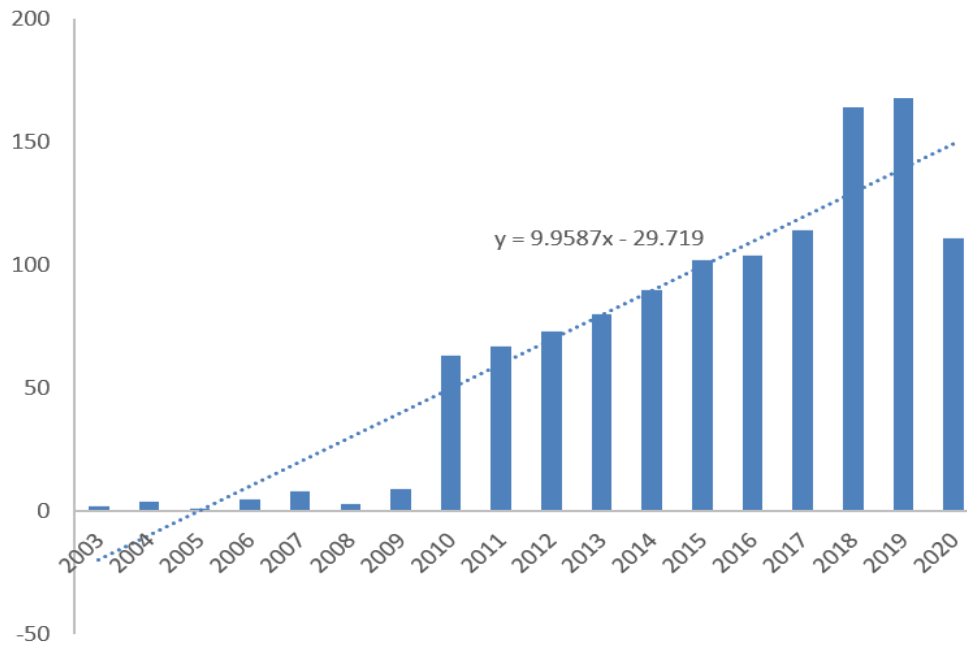**Figure 3:** Systematic literature review

Figure 4: Number of related publications published every year

Text analytics extracts quality and actionable insights from textual data by using various Natural Language Processing (NLP) techniques to aid the end-user's understanding and decisions in a situation. It often involves gathering data, pre-processing data, data representation, and finding valuable knowledge and insights from the data. Data can be gathered in several ways such as extractions from social media as well as using statistical and real data that has been collected over time. In this study, publication data from several databases were collected and processed before analysis. Data pre-processing depends on the need of the user, which could include the removal of stop words and tokenization. Once the data was processed, it was modeled into a numeric vector form; such as Bag-of-Words (BOW) and word2vec. Finally, data visualization was used to gain knowledge from the data.

Text analytics has been used in various works to understand human behavior; such as in hospitality management (Ting et al., 2017) and disaster response (Lee et al., 2018) as well as in user profiling (Alekseev & Nikolenko, 2017; Ahmad et al., 2019). In this study, text analytics was used to explore and understand human perception and awareness of road safety by gathering data from publications in several popular databases. Numerous previous studies have utilized surveys and questionnaires to understand the behaviors of road users (Cordellieri et al., 2016; Lee & Han, 2018); however, this traditional method faces drawbacks such as: (1) limited number of participants in comparison to population, (2) time-consuming and costly, and (3) biased feedback from personal experiences that can be manipulated. Text analytics, on the other hand, allows for this data to be gathered with: (1) the ease of accessing vast quantities of data for analysis, (2) obtaining natural data on human behaviors, and (3) convenience of conducting a longitudinal study by capitalizing on an abundance of timely social media data.

J̇SAEM

3.0 RESULTS AND DISCUSSION

The text analysis was performed after data gathering and used to investigate the most common phrases used in publications on the awareness of road and traffic safety. A list of texts was collected from the titles, keywords, and abstracts of relevant publications. Several words were excluded to avoid biases in the process while stop words were also excluded as they are the most common words used in a language. Three different sources of words from publications were used: (1) the title column, (2) the keywords column, and (3) the abstract column. The total number of words used in the title was 1610.

Figure 5 shows a bar graph of the frequencies of words used in the titles of all publications. The word "drivers", "fatigue", "injuries", "crashes", and "risk" provided vital clues as to the topic discussed in the publication. Figure 6 elucidates the connection between two words to provide a better understanding of the single words used in Figure 6. For example, "driver fatigue" is the third-highest combination of words used in road and traffic safety publications. Driving fatigue; a feeling of drowsiness due to an extended driving period, monotonous road conditions, adverse climatologically, or a drivers' individual characteristics; can either be direct or contributing factors to road accidents (Abdullah & Von, 2011). In Malaysia, a number of traffic collisions occur in the early hours of the morning often resulting in serious casualties. These crashes have, in part, been partially attributed to fatigue and sleepiness (Mohamed et al., 2012; Wahida et al., 2013). In Malaysia, as with most countries around the world, drivers who operate commercial vehicles; such as taxis, buses, or trucks; normally work on shift-based rotations that require them to maintain irregular and abnormal sleeping hours that, more often than not, leave them feeling fatigued and sleepy. This exhaustion combined with drivers suffering from Obstructive Sleep Apnoea (OSA) who are required to drive long distances or during irregular hours only further compounds the risk of traffic-related accidents.

The term "seat belt" also appeared in the list which indicates its importance in the topic of public awareness. Seat belts are one of the most efficient and inexpensive in-vehicle safety devices capable of reducing the risk of death or serious injury in a crash by almost 50% for drivers and front-seat passengers and by almost 25% for rear-seat passengers (Jawi et al., 2016). However, the strict enforcement of seat belt usage is only a short-term solution that requires technological intervention in the long term. It has been suggested that the relevant agencies may make it compulsory for all Malaysian carmakers to devise a system wherein the vehicle will not start if seat belts are not fastened in the driver and passenger seats (Sambasivam et al., 2014).

The total number of words used in the abstracts columns was 24,500 words. Figure 7 presents the most frequently used words from the abstract column using the 1-gram and 2-gram models. Word clouds were used to visualize the frequency of word sequences. A powerful visualization technique in text analysis, it was used to illustrate the most frequently used single or two-word sequences used in the publication titles. The most popular single word sequences were "driving", "injuries", "fatigue", and "health" while "public health", "seat belt", "while driving", and "traffic injuries/accidents" were at the top of the list of two-word sequences. As these words appeared in both single and two-word word clouds, the public considers them important parameters in traffic safety. A visualization of the words used in abstracts helps researchers to find important topics to consider for further research. The words "motor vehicle", visualized in Figure 7(b), also alludes to the importance of information pertaining to motorcycle-related injuries in relation to traffic safety. Statistics on motorcycle-related injuries

in Malaysia, involving either single or multiple vehicles, could provide valuable insight with which to solve traffic safety issues for both motorcyclists and other road users as motorcycles account for 47% of all registered vehicles in the country while motorcyclists account for 59% of all reported accident-related fatalities (Manan, 2014).
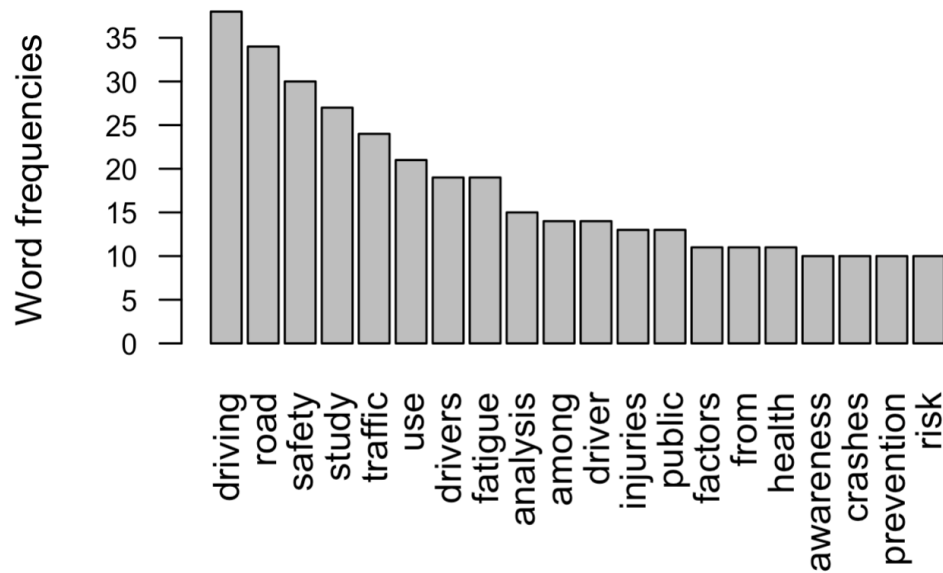

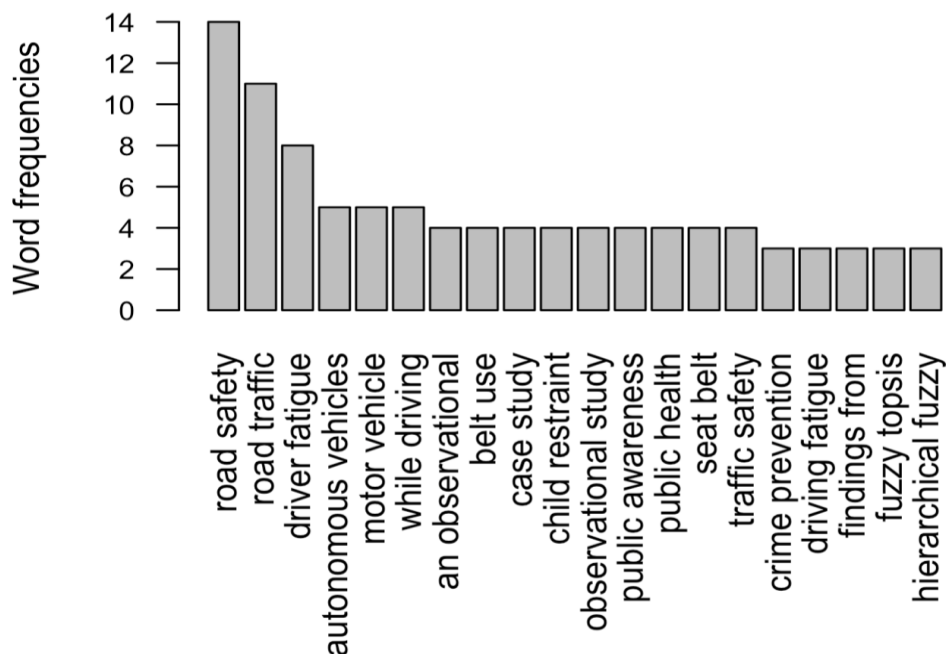**Figure 5:** A frequency of one word used in the title

Figure 6: A frequency of two words used in the title

Human factors, followed by vehicle factors, were found to be the main causes of accidents while illegal driving and jaywalking (crossing or walking on the road without regard for approaching traffic) were the main causes of car accidents. Therefore, the strengthening of public education on traffic safety (especially for the floating population), providing drivers with rigorous training of traffic rules and regulations, enhancing public awareness of road safety as well as improving road traffic management and control measures were the main measures suggested to prevent and control traffic-related injuries. These interventions as well as education regarding the risks associated with engaging in secondary activities while driving, law enforcement, tougher legislation, periodic assessment, raising public awareness as well as attracting political and social support should all be integrated into the WHO's Safe Community network in cities.

(a)   (b)

Figure 7: Keywords from abstracts of related publications shown in (a) 1-gram word cloud and (b) 2-gram word cloud models

Figure 8 shows the plethora of issues being discussed in two different databases. Figure 8(a) displays the 1-gram word cloud of the Scopus abstract and citation database. Here, the frequency of words used by researchers focused more on human factors; such as driver behavior, human injuries, health, fatigue, children, and pedestrians. Figure 8(b) shows the 1-gram word cloud of the Institute of Electrical and Electronics Engineers (IEEE) digital library research database where different words relating to technology and method and performance; such as vehicle, communication, system, and information: were used more frequently by researchers to discuss road and traffic safety awareness in detail. This is because IEEE articles focus on new technologies, technological enhancements, and methodologies to address problems whereas Scopus articles are more general and focus more on the social sciences to discuss this topic. Therefore, the visualization approach enabled us to view the issue from two different perspectives that led to a better understanding of public awareness of road and traffic safety. These two perspectives were then supported by the frequency of two-word sequences used in both databases (Figure 9 and Figure 10).
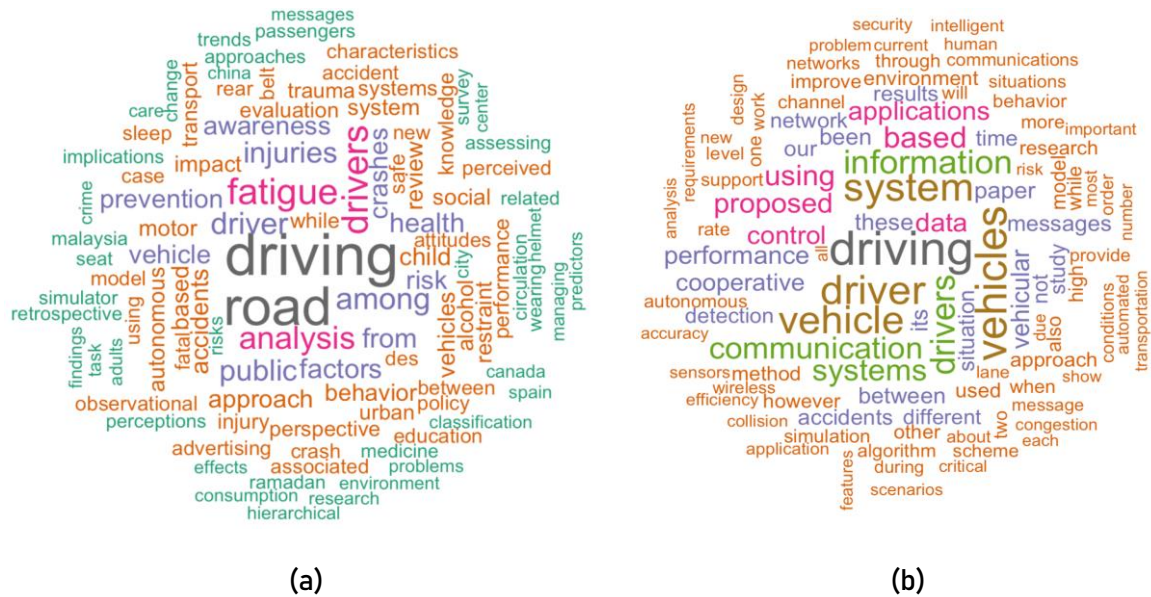
(a)   (b)

Figure 8: 1-gram word cloud from (a) Scopus and (b) IEEE
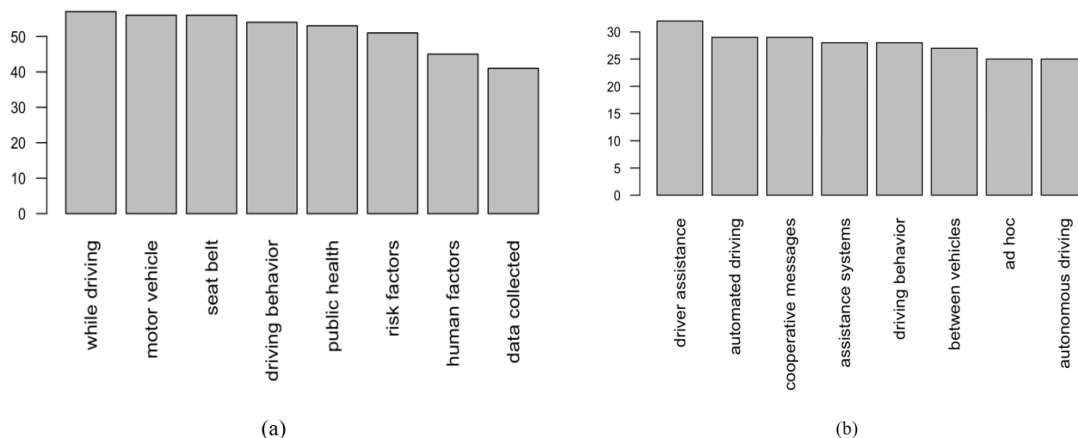

Figure 9: Frequency of two-word sequences used in abstracts on (a) Scopus and (b) IEEE

Technologies that are being implemented to improve road and traffic safety are visualized in Figure 11. An example of the technologies used is driver impairment detection using Dynamic Bayesian Network (McDonald et al., 2018). Table 2 shows the frequency of three-word sequences used in publications to illustrate a group of road safety issues. The top three most frequently used words are related to "crashes" or "accidents" as they are a major concern for the public. Despite the success of various engineering, education, and enforcement measures, fatalities and injuries from traffic collisions remain a major global problem. In order to successfully address this growing issue, it has been suggested that road users may require a fundamental reform of their Traffic Safety Culture (TSC). As such, measuring and understanding traffic safety culture has gained popularity. Studies have revealed that most traffic collisions are caused by driver-associated factors as road stakeholders often overlook the importance of traffic safety resulting in hazardous crashes that cause devastation not only to themselves but to others. This growing interest in TSC is a key factor in managing and sustaining safe roadway transportation systems, especially to reduce traffic fatalities and serious injuries. However, the theories, terminologies, and methodologies of addressing TSC

comes from the human and social science disciplines that are not typically involved in traditional traffic safety, engineering, or other behavioral change agencies; such as the department of transportation, driver licensing, motor vehicle records, etc. Therefore, a fundamental lack of shared language and understanding of TSC limits the ability of agencies to explore this topic and engage new stakeholders.

(a)   (b)

Figure 10: 2-gram word cloud from (a) Scopus and (b) IEEE

Figure 11: 2-gram word cloud of technologies that are being implemented to improve road and traffic safety

113

Table 2: The frequency of three-word sequences used in Scopus

Phrases	Frequency
motor vehicle crashes	10
road traffic accidents	10
road traffic crashes	10
motor vehicle collisions	8
crash avoidance maneuvers	7
road traffic injuries	7
fatally injured pedestrians	6
motor vehicle risk perception	6
rear seatbelt wearing	6
texting while driving	6
environmental noise pollution	5
fatal crash involvement	5
injured pedestrians bicyclists	5
pedestrian crash risk	5

4.0 DISCUSSION AND CONCLUSION

In recent decades, the evolution of public awareness has influenced the way that road and traffic safety initiatives are designed and implemented. Communication and education through the distribution of road safety information are some of the most important elements of all road safety strategies. Therefore, it is imperative that road safety communicators not only stay abreast with the latest technologies but utilize the newest communication tools to convey road safety messages as effectively as possible. As the media industry has undergone significant changes over the past few years, it has provided a voice for persons other than professional journalists. Platforms, such as social media, open a whole new world of communication that can assist in changing driver behaviors and enhancing road safety.

Our successful utilization of text analytics demonstrates its ability to effectively generate and codify rich and quality knowledge for researchers in the field to better understand current research trends. However, it is not without its limitations that could serve as an important starting point for further research. In our exploratory study, text analytics was applied only to articles published in several databases over a ten-year period. Therefore, future studies could apply text analytics to a longer period to examine trends and to consider more resources to develop the robustness of their findings. For instance, different sources of unstructured data such as websites, social media feeds, and others resources available on the Web have the potential to generate useful insights.

JSAEM

ACKNOWLEDGEMENTS

REFERENCES

Abdullah, A. D. N. M., & Von, H. L. (2011). Factors of fatigue and bus accident. In International Conference on Innovativation, Management and Service, 317-321.

Ahmad, S. S. S., Osman, A. N. B. S., & Basiron, H. (2019). The impact of social media on human interaction in an organisation based on real-time social media data. International Journal of Data Science, 4(3), 260-271.

Alekseev, A., & Nikolenko, S. (2017). Word embeddings for user profiling in online social networks. Computacion y Sistemas, 21(2), 203–226.

Cordellieri, P., Baralla, F., Ferlazzo, F., Sgalla, R., Piccardi, L., & Giannini, A. M. (2016). Gender effects in young road users on road safety attitudes, behaviors and risk perception. Frontiers in Psychology, 7, 1412.

Darma, Y. (2017). A time series analysis of road traffic fatalities in Malaysia. Faculty of Engineering, University of Malaya.

FIA Foundation (2006). Make roads safe – A new priority for sustainable development. Retrieved from http://www.who.int/management/programme/health_promotion/MakeRoadsSafe.pdf

Hue, D. T., Brennan, L., Parker, L., & Florian, M. (2015). But I AM normal: Safe? Driving in Vietnam. Journal of Social Marketing, 5(2), 105-124.

Jawi, Z. M., & Kassim, K. A. A. (2013). 3-5-2: How does NCAP for ASEAN help the region's road safety index? In Proceedings of the Southeast Asia Safer Mobility Symposium, Melaka, Malaysia.

Jawi, Z. M., Isa, M. H. M., Mohamed, N., Awang, A., & Osman, M. R. (2016). A systemic analysis of the usage of safety items among Malaysian private vehicle users. Journal of Mechanical Engineering and Sciences, 10(3), 2262-2274.

Kitamura, Y., Hayashi, M., & Yagi, E. (2018). Traffic problems in Southeast Asia featuring the case of Cambodia's traffic accidents involving motorcycles. IATSS Research, 42(4), 163-170.

Lee, A. S. H., Yusoff, Z., Zainol, Z., & Pillai, V. (2018). Know your hotels well! an online review analysis using text analytics. International Journal of Engineering & Technology, 7(4.31), 341–347.

Lee, H., & Han, S. (2018). Characteristics of road users' safety behavior and attitude in Korea : Findings from ESRA survey. 18th International Conference Road Safety on Five Continents, 1(12), 1–12.

Lum, M. (2019). We have the third highest death rate from road accidents. The Star. Retrieved from https://www.thestar.com.my/lifestyle/health/2019/05/14/we-have-the-third-highest-death-rate-from-road-accidents

Manan, M. M. A. (2014). Factors associated with motorcyclists' safety at access points along primary roads in Malaysia. Doctoral dissertation, Lund University.

McDonald, A. D., Lee, J. D., Schwarz, C., & Brown, T. L. (2018). A contextual and temporal algorithm for driver drowsiness detection. Accident Analysis & Prevention, 113, 25-37.

Mohamed, N., Mohd-Yusoff, M. F., Othman, I., Zulkipli, Z. H., Osman, M. R., & Voon, W. S. (2012). Fatigue-related crashes involving express buses in Malaysia: Will the proposed policy of banning the early-hour operation reduce fatigue-related crashes and benefit overall road safety? Accident Analysis & Prevention, 45, 45-49.

MOT (2019). Transport Statistics Malaysia 2018. Ministry of Transport, Malaysia.

Sambasivam, S., Karuppiah, K., Shamsul Bahri, M. T., Kulanthayan, S., & Sadeghi, H. N. (2014). Non-compliance of Malaysia motor vehicles (safety seatbelts) rules 1978. Iranian Journal of Public Health, 43(3), 1-6.

Ting, P.-J. L., Chen, S.-L., Chen, H., & Fang, W.-C. (2017). Using big data and text analytics to understand how customer experiences posted on Yelp.com impact the hospitality industry. Contemporary Management Research, 13(2), 107-130.

UN (2016). UN Study: Transport for Sustainable Development – The case of inland transport.

UNECE (2019). Road Safety for All. Sustainable Development Goals (SDG).

Wah, T. W. (2015). Miros statistics say human error causes 80% of traffic accidents. The Sun Daily.

Wahida, A. B., Ilhamah, O., Suffian, A. M., Aimi, M. F., Norlen, M., & Wong, S. V. (2013). Obstructive sleep apnea among commercial vehicle drivers in Malaysia: Issues and initiatives. Health, 05(08), 80-86.

WHO (2011). Global launch: decade of action for road safety 2011-2020. World Health Organization, No. WHO/NMH/VIP11. 08.

WHO (2015). Road safety: Basic facts. World Health Organization.

WHO (2018). Global status report on road safety 2018. World Health Organization.