KNOWLEDGE TRANSFER ABOUT EARTHQUAKE DISASTER MITIGATION TO CHILDREN THROUGH TF-IDF

Maria Umran[*], Hafiz Mohd. Sarim[**]
[*]*Faculty of Information Science and Technology, Universiti Kebangsaan Malaysia, Selangor, Malaysia, maria.alkibt@gmail.com*
[**]*Faculty of Information Science and Technology, Universiti Kebangsaan Malaysia, Selangor, Malaysia, hms@ukm.edu.my*

Email Correspondence : maria.alkibt@gmail.com

Received : July 1, 2020 Accepted : December 8, 2020 Published : December 31, 2020

Abstract: Past observations during a disaster identify that when children are separated from parents, they suffer due to the inability to comprehend disaster mitigation concepts. This study proposes a process from the existing framework K-Nearest Neighbor (KNN) and Term Frequency - Inverse Document Frequency (TF-IDF) for extracting a large body of knowledge in the form of documents into simple words. Those simple words can be arranged into contextual lyrics utilizing an Artificial Intelligence lyrics generator and then orchestrated into a song using a music generator. The piece, which is the output of the proposed process, is utilized to transfer the knowledge about earthquake disaster mitigation to children. A quantitative analysis of questionnaires on students aged 9-10 in Banda Aceh shows the song's highly significant effect in transferring the knowledge about earthquake disaster mitigation to children.

Keywords: TF-IDF; disaster mitigation; knowledge transfer

Recommended APA Citation :
Umran, M., & Sarim, H. M. (2020). Knowledge Transfer About Earthquake Disaster Mitigation To Children Through TF-IDF. *Elkawnie*, *6*(2), 165-177. https://doi.org/10.22373/ekw.v6i2.7281

Introduction

As one of the most unpredictable natural disasters, earthquakes have been known to contribute to considerable loss and destruction. On June 5, 2015, a strong earthquake at the moment magnitude of 6.0 on the Richter scale in Sabah, Malaysia, killed six Singaporean children from the Tanjong Katong Primary School (BBC News, 2015). The Guardian (2011) reports that an earthquake and its resulting tsunami devastated Japan, and may have displaced as many as 100,000 children when many were at schools or kindergarten in the mid-afternoon. Due to the difficulty in predicting earthquakes, there is no guarantee that children will always be along with their parents when a disaster occurs. Therefore, to prevent injury or loss of life, children need to gain some knowledge about disaster mitigation to equip themselves with the survival skills to survive earthquakes.

However, transferring knowledge to children is not an easy task due to their limited absorptive capacity. There are still problems regarding children's behaviour and attitude on natural disasters and practical knowledge, even though they have already learned about natural disasters and mitigations at school for years. This happened due to ineffective teaching technique, which depends much on pictures and textbooks as the teaching method (Tuswadi & Hayashi, 2014). Consequently, a newly generated medium, such as songs, can be considered as a teaching medium for children. According to Walton (2016), children seemed to learn new phrases more efficiently and retain words longer when they are composed within jingles, which are brief songs. A well-known example of songs as generational knowledge is the "Duck and Cover" animated film. It was intended to be a civil defence counter for children in a nuclear bomb attack during the Cold War in the 1950s and 1960s (E. Matthew, Jr., 2011).

Another study finding shows that low scores in children's questionnaires about disasters may be attributed to teaching materials that are inconsistent with the key messages provided by the authority (Amri et al., 2017). This study's proposed process develops a mechanism that can be modified based on the needs in conveying the key messages from stakeholders to children. On the other hand, children with a lack of literacy and different education levels limit their perceptions of a potential emergency and the need to prepare for it.

Transferring knowledge about disaster can also be conducted through games. A flood protection computer game for disaster education motivates students to learn about typhoons and flood mitigation in Taiwan effectively (Tsai et al., 2011). However, one of the difficulties in training children is the significant difference between the education level in the big cities and small towns or villages due to the illiteracy rate (Izadkhah & Hosseini 2005). In some cases, the children in villages have limited access to electricity or computers. Therefore, a new medium such as songs will erase the boundaries between literate and less literate, privileged and underprivileged children, as a means to transfer equal knowledge about disaster mitigations in both cities and remote areas.

A proposed process is required to digest a large body of knowledge in documents by utilizing Knowledge Management Tools to produce simple words and phrases. Those words are then arranged into lyrics and orchestrated into documented songs. Knowledge extraction can be performed through text classification with the TF-IDF algorithm, which shows the high accuracy of 98.3% in Bahasa Indonesia (Hakim et al., 2015). TF-IDF can also be utilized as a word extraction method to detect the most related keywords from documents (Doen et al., 2014). These past researches become an imagination to perform something similar in extracting the knowledge about disaster mitigation, and then convey it to children in the form of 'Combination' knowledge conversion. Nonaka and Takeuchi (1995) define 'Combination' as knowledge conversion of explicit

knowledge, such as documents, to other direct experiences, such as reports, spreadsheets, books, or music notes.

Those past researches, on the other hand, only accomplish knowledge extractions. Hence, a proposed process is required to modify the knowledge conversion of 'Combination' to a nouveau knowledge transfer mechanism. This mechanism combines complete ad hoc mechanisms of knowledge extraction, lyrics creation, and song orchestration, considering none of Knowledge Management Tools aims to integrate all instruments from start to finish, specifically in earthquake disaster mitigation. Through a knowledge extraction and knowledge transform mechanism, a newly generational knowledge can be created in songs.

To arrange simple words into contextual lyrics, DeepBeat as a lyrics creation tools is required in the process. It is based on two machine-learning techniques. Those are a deep neural network model and the RankSVM (Support Vector Machine). It shows 82% the accuracy of the distinguished real next line from a randomly selected lyric (Malmi et al., 2015). By listening to the song, children as the recipients can easily retain the intended lessons and be proactive during disasters to minimize injuries, prevent trauma, and reduce calamities.

Aims of the Study

The general objective is to use Knowledge Management System to transfer knowledge from experts' works in earthquake disaster mitigation to children through songs. Meanwhile, the study's specific objectives are; 1) to propose a process using the existing Knowledge Management Systems framework to extract a large body of knowledge into simple and easily understandable words and develop a song's rhythmic structure from the extracted words. 2) to evaluate the effectiveness of the newly generated song in transferring disaster mitigation knowledge to children.

Literature Review

One song in particular that has been attributed to the survival of thousands of people's lives in Simeulue Island, Aceh, during the earthquake and tsunami in 2004 is the story of Smong (meaning "tsunami" in the Devayan language). It has been proven that victims in other parts of Aceh have larger casualty ratios, with the death toll reaching more than 200,000 people out of an original population of 4,193,000 (1:20 ratio). Meanwhile, in Simeule island, the victims were 7 people out of 78,000 in total population with a 1:11,000 ratio (Syafwina, 2014). According to Mcclain Opiyo (2015), music or songs play a significant role in storing information for current generations. The future generation will have a chance to know about it.

Converting a large body of knowledge is feasible by using knowledge extraction. Saitoh et al. (2015) propose a feature selection based on onomatopoeia

for the knowledge construction model by focusing on feature selection of text mining. Another knowledge extraction framework utilized is text classifiers, where Jie et al. (2015) construct the knowledge extraction framework. A large text's segmentation process on a document into several clusters and knowledge can be identified based on text classifier, which is constructed with each form's significant elements (Chen et al., 2018).

Experiments and Evaluation

Proposing a Process from KNN & TF-IDF Framework

The flow from knowledge extraction, lyrics generation, and song orchestration were patched in a proposed process, which creates a newly generated medium. This process implements the KNN & TF-IDF (Term Frequency - Inverse Document Frequency) framework (Trstenjak et al. 2014), proposed to assist the measurement and classification of documents similarity based on the needed text sample. Trstenjak et al. (2014) also employ the K-Nearest Neighbor (KNN) as the simplest machine learning algorithms. Meanwhile, the process for this study is implemented in "RapidMiner" as a Knowledge Management Tool, "DeepBeat" as lyrics generator, "WolframTones" as songs orchestrator, and "Sibelius" as music note writer. The presentation of the proposed process can be displayed in Figure 1:

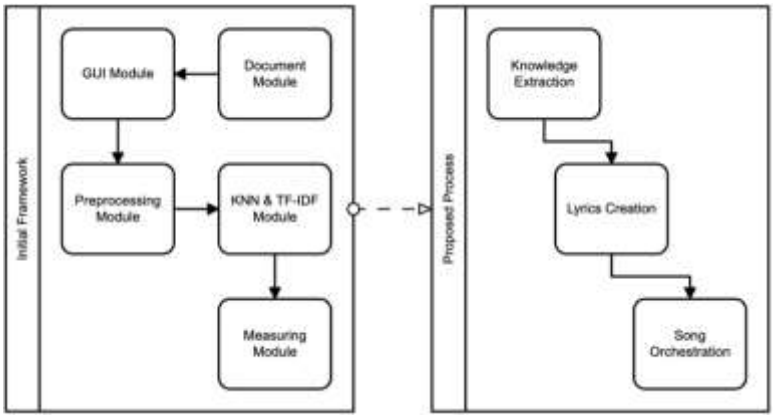

Figure 1. A Proposed Process to create a Newly Generated Medium from Initial Framework by Trstenjak et al., (2014)

Identifying Significant Words

To digest a large body of knowledge, a knowledge extraction process will be deployed by using a Knowledge Management Tool through these steps, which are; 1) selecting relevant documents, 2) Tokenizing documents, 3) Transforming cases, 4) Filtering stop words, 5) Generating n-Grams in terms, 6) Filtering tokens by length. Knowledge extraction uses term frequency and TF-IDF. There are three documents about earthquake disaster mitigation, which will be extracted, are; 1) "Great Shake Out Earthquake Drills", published by Earthquake Country Alliance, United States of America 2) "What to Do in an Earthquake", released by Victoria

State Emergency Service Australia, 3) "Earthquake", published by Department of Homeland Security, United States of America.

After extracting those documents, the highest frequent until the least frequent phrases appeared on the result, where it was limited to 8 total occurrences. Therefore, there are 23 important words as a result of the knowledge extraction technique. The highest occurrence word is "earthquake" with a total occurrence of 41. Meanwhile, the least occurrence word is "emergency" with a whole occurrence of 8. It can be assumed that those words or phrases are the most relevant words regarding earthquake disaster mitigation. These contextual phrases were utilized as inputs to arrange lyrics in the next step of this research experiment. Significant words are displayed in Figure 2:

Earthquake	Cover	Hold	Drop	Protect	Stay
Shaking	Shelter	Ground	Head	Buildings	Move
Table	Glass	Situations	Stop	Doorway	Nearby
Neck	Debris	Emergency	Crawl	Falling	

Figure 2. Significant Words

Rearranging Significant Words into Lyrics

After performing the knowledge extraction process, the significant words will be put into a lyric generator. In this study, DeepBeat will be deployed as a musical generator. This generator helps users create a whole new song verse by combining rhymes from existing songs or selecting some options to choose which rhyme fits the best since the music must be contextual. Arranging phrases or words into lyrics was not that easy since the lyrics' flows and common sense must be considered. As it is concerned that the final result will be delivered to children, the phrases must be simple but still contextual. In this research, the arrangement of the lyrics can be seen below:

"When you feel ground shaking, drop, cover, hold on
Drop to your hands and knees, protect your head and neck
Crawl and hold on to sturdy tables stay until earthquake stops
Do not move into doorway cause building falling down
Do not run outside cause debris and glass can hurt
In emergency situation find shelter nearby and be aware"

Orchestrating Lyrics into a Song

After using DeepBeat to generate the lyrics, orchestration software such as WolframTones creates a musical composition. These music notes can be played with a piano as a musical instrument or with any other device. To combine lyrics and design into a readable form, anyone can preserve and use the composition to play instruments and sing the song. A software named Sibelius has been utilized

to achieve this aim. This platform generates, plays, and print sheet music for practising notes. The music notes and lyrics are displayed in Figure 3:

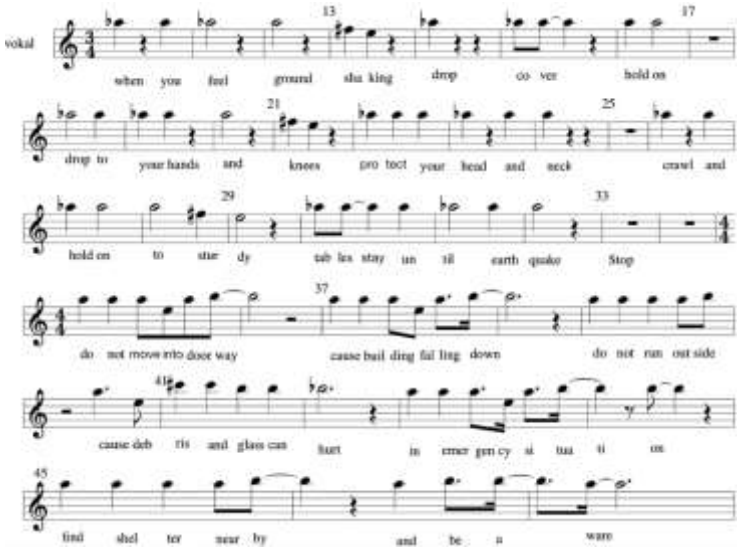

Figure 3. A Song with Lyrics Constructed from the Proposed Process

Evaluating the Effectiveness of the Newly Generated Medium

To evaluate the effectiveness of the newly generated medium and whether it is understandable by children, two types of close-ended questionnaires were distributed to 3rd-year primary school students aged 9-10 at Bunga Matahari International School in April 2019 in Banda Aceh. The school was chosen using purposive sampling with criteria focused on; 1) location, Banda Aceh experienced a massive impact of the earthquake in 2004. 2) Language, teaching, and learning process at the school is conducted mostly in English. This study's output is naturally in English only because the sources are in English only, and the "Language Translation" component is not available in the proposed process. 3rd-year students were selected to answer the questionnaires because their English ability is sufficient to answer the questions.

The total population is 41 students, and the number of respondents is 27 students. The number of respondents was obtained by employing a sample size formula, where the margin of error is 0.12, the confidence level is 0.95, and the response distribution is 0.5. The gender of respondents is ignored in determining the respondents in this study. Before disseminating the questionnaires, they have been tested on 5 students aged 7 until 11. The questionnaires consist of two types of questions. Those are; BEFORE questionnaire and AFTER questionnaire. They have the same 10 questions about the song's disaster mitigation lyrics to improve students' knowledge in earthquake disaster mitigation. The Validity and Reliability Tests on the questionnaires can be seen in Table 1:

a. Validity Test

Table 1. Correlation Output

Question	r Calculation	r Table	Result
1	0.881	0.632	Valid
2	0.852	0.632	Valid
3	0.852	0.632	Valid
4	0.852	0.632	Valid
5	0.881	0.632	Valid
6	0.881	0.632	Valid
7	0.852	0.632	Valid
8	0.881	0.632	Valid
9	0.852	0.632	Valid
10	0.852	0.632	Valid

Validity Decision
If r Calculation > r Table with Sig. (2-tailed) of 0.05, then the instrument is declared valid.
If r Calculation < r Table with Sig. (2-tailed) of 0.05, then the instrument is declared invalid.

Result Interpretation

Based on the result of Pearson Product Moment Correlations for questions 1 to 10, the correlation output is depicted in Table 1. According to Table 1, the correlation value of r Calculation for questions 1 to 10 is more significant compared to the r Table. Therefore, it can be concluded that all instruments in the questionnaire are valid.

b. Reliability Test

Table 2. Reliability Statistics

Cronbach's Alpha	N of Items
0.963	10

Reliability Decision
If Cronbach's alpha > 0.6, then the instruments are consistent or reliable.
If Cronbach's alpha < 0.6, then the instruments are inconsistent or unreliable.

Based on Cronbach's alpha test result to examine the reliability and essential decision-making for data reliability, the obtained information shows that the instruments are reliable or consistent since the value of Cronbach's alpha is 0.963, which is greater than 0.6.

Initially, the students answered the BEFORE questionnaires prior to listening to the song. After that, the students heard the song three times then answered ten questions on the AFTER questionnaire based on their perspective

towards the song. All answers were then examined based on quantitative analysis to discover the result. Based on a rough calculation on Microsoft Excel, 16 out of 27 students got score escalation. It means that there is a positive knowledge improvement for 59,25% in this group study. There is also an improvement of the score up to 9 to 10 correct answers from 5 to 6 correct answers. The result of the survey on Microsoft Excel is depicted in Figure 4:

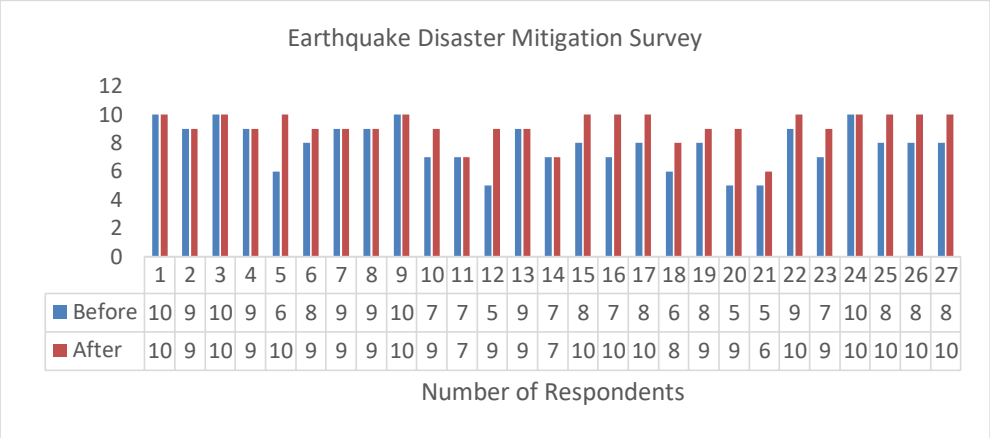

Figure 4. Chart of Disaster Mitigation Survey

To show the result in a more in-depth insight of data analysis in measuring the significance of the result, SPSS has been deployed as follows;

a. Data
   1) Data from BEFORE questionnaires
   2) Data from AFTER questionnaires

b. Demand Analysis
   1) Normality test by utilizing Shapiro-Wilk Test
   2) Normality test by utilizing Kolmogorov-Smirnov Test

Table 3. Normality Test

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	**df**	**Sig.**	**Statistic**	**df**	**Sig.**
Before	0.168	27	0.049	0.921	27	0.043
After	0.296	27	0.000	0.746	27	0.000

a. Lilliefors Significance Correction
b. df (degree of freedom)
c. Sig. is Significance

c. Normality Hypothesis

H_0: $\mu_1 = \mu_2$ or data is distributed normally or normal population distribution.

$H_{1:}$ $\mu_1 \neq \mu_2$ or data is not distributed normally, or population distribution is not normal.

d. Theory of Decision

If Sig. > α (alpha), it means that H_0 is accepted.

If Sig. < α (alpha) means that H_0 is rejected or there is not enough evidence to accept H_0.

e. Result for Normality Test

In this study, α (alpha) value or probability is 5 %. Based on received information from data, the number of respondents is 27 samples, relatively only a few in numbers. Therefore, it is recommended to apply the Shapiro-Wilk Test compared to the Kolmogorov-Smirnov Test for the two data. Based on the normality test for 27 respondents by exercising Shapiro-Wilk Test, BEFORE questionnaires show a count Statistic 0.921 with Significance value (Sig.) 0.043. Since Significance value is smaller than alpha or the mentioned probability 0.05 (Sig. = 0.043 < α = 0.05), according to the decision theory for H_0 is rejected. So, the normality test's final result on the BEFORE questionnaire shows that data is not distributed normally.

Based on the normality test on respondents for AFTER questionnaires, utilizing the Shapiro-Wilk Test gathered information displays that count Statistic is 0.746 with Significance value (Sig.) 0.000. As Significance value is smaller than alpha or probability 0.05 (Sig. = 0.000 < α = 0.05), consequently the decision theory for H_0 is rejected. Hence, the outcome for the normality test on AFTER questionnaires demonstrates that data is not distributed normally.

f. Difference Test by Employing Wilcoxon Signed Ranks Test

Wilcoxon Signed Ranks Test is implemented since both data are not distributed normally, and the samples are related to each other. The BEFORE questionnaires present it has exact questions as AFTER questionnaires. The result of the Wilcoxon Signed Ranks Test on BEFORE questionnaires and AFTER questionnaires can be perceived in Table 4:

Table 4. Statistics Test[a]

	Result - Time
Asymp. Sig. (2-tailed)	0.000

a. Wilcoxon Signed Ranks Test

b. Based on negative ranks
c. Asymp. Sig. is Asymptotic Significance

For the Wilcoxon Signed Ranks Test, the α (alpha) value or probability is also 5%, and this value is the probability of rejecting the null hypothesis when the null hypothesis is correct. The alpha level should be determined before conducting the hypothesis.

g. Hypothesis
   H_0: there is no difference for BEFORE and AFTER questionnaires.
   H_1: there is a difference for BEFORE and AFTER questionnaires.
h. Basic of Decision
   If count Statistic > probability, then H_0 is accepted.
   If count Statistic < probability, then H_0 is rejected.

In decision making for the Wilcoxon Signed Ranks Test, it is visible that Asymp. Sig. is 0.000. Meanwhile, the value of Asymp. Sig. is smaller than probability (0.000 > 0.05), it can be concluded that H_0 is rejected or there is a difference for BEFORE and AFTER questionnaires. Asymp. Sig. 0.000 also shows a significant difference or improvement of score escalation on AFTER questionnaires. The song's effectiveness can also be observed from the escalation of the average value from BEFORE and AFTER questionnaires. The total score of 27 students for BEFORE questionnaires is 212, so the average value is 7.85. Meanwhile, the total score for AFTER questionnaires is 247, and the average value is 9.15. It reflects the highly significant effect of the song as a learning medium in transferring and improving the knowledge about earthquake disaster mitigation. Therefore, this can be appreciated as an important contribution to disaster mitigation, especially in earthquake disaster mitigation.

Results and Discussion

This study contributes to a proactive measure for children in mitigating disasters, especially earthquakes. This study also gives a constructive and applied contribution to the research of Knowledge Management and Knowledge Management Systems. Users often deploy Knowledge Management Systems to extract knowledge, but they never utilized it to reform learning. This possibly provides a new role for Knowledge Management Tools to repackage the extracted knowledge to targeted audiences. Another ultimate impact of this study is reducing the number of earthquake victims through the spread of generational knowledge contained in songs.

According to children's evaluation as end-users, the song as a newly generated medium has significant effectiveness in transferring knowledge about earthquakes, especially earthquake disaster mitigation to children. By

implementing the process and using the song as a newly generated medium, it is believed that children remember the intended lesson much more easily. The song also helps them to survive during an earthquake, not only to prevent physical injuries but also to save their lives and reduce more significant losses in their surroundings.

Based on this study, alternative solutions in the future for knowledge extraction are as follows; a) Variety of primary sources of information for knowledge extraction. In the end, website contents, data from databases or social media can be one of the main sources as a large body of knowledge containing substantial knowledge about disaster mitigation. Another solution is b) Availability of Information. Nevertheless, there are disasters or accidents, which are not so common or do not happen frequently but still need to be measured. For instance, liquefaction is a disaster that sometimes follows after an earthquake. It does not have well-documented information in regards to its mitigation. On the other hand, industrial accidents also impact children negatively, such as industrial gas leaks in Johor (The Straits Times, 2019), but unfortunately the information or source is harder to define since it does not happen regularly and the root cause lacks commonality. Therefore, this study's work needs to adapt to well documented or less documented disasters in coming times.

Conclusion

Overall, this study has achieved research purposes in designing a proposed process, which can be implemented and modified by potential stakeholders and users. Potential stakeholders include governments, ministries, United Nations, Non-Governmental Organizations, mass media, schools, teachers, parents, etc. Furthermore, in an attempt to improve Knowledge Management, this study provides a new avenue for using this technology. The technique has been utilized in this study, can also be used for other disaster mitigations, natural or human-caused disasters, such as fire, industrial accidents, act of terrorism, shootings, and mass violence incidents.

Information about earthquake disaster mitigation can be readily acquired compared to information about smaller disasters or less frequent disasters. These kinds of disasters are not well documented and even less understood for people, for instance, the outbreak of infectious diseases, community upheaval, and so on. This kind of event or disaster needs to be considered and documented to give new research possibilities in disaster mitigation and the Knowledge Management field in the future.

Acknowledgement

This research was sponsored by the GGP-2017-076 research grant from the Centre for Research Instrumentation and Management (CRIM), Universiti Kebangsaan Malaysia.

References

Amri, A., Bird, D. K., Ronan, K., Haynes, K., & Towers, B. (2017). Disaster Risk Reduction Education in Indonesia: Challenges and Recommendations for Scaling Up. https://doi.org/10.5194/nhess-17-595-2017

BBC News. (n.d.). Malaysia Quake: Singapore and Sabah Mourn Victims - BBC News. Retrieved October 27, 2018, from https://www.bbc.com/news/world-asia-33039206

Chen, J., Zhang, C., & Niu, Z. (2018). A Two-Step Resume Information Extraction Algorithm. https://doi.org/10.1155/2018/5761287

Doen, Y., Murata, M., Otake, R., Tokuhisa, M., & Ma, Q. (2014). Construction of Concept Network from Large Numbers of Texts for Information Examination Using TF-IDF and Deletion of Unrelated Words. https://doi.org/10.1109/SCIS-ISIS.2014.7044701

Hakim, A. A., Erwin, A., Eng, K. I., Galinium, M., & Muliady, W. (2015). Automated Document Classification for News Article in Bahasa Indonesia Based on Term Frequency Inverse Document Frequency (TF-IDF) Approach. https://doi.org/10.1109/ICITEED.2014.7007894

Izadkhah, Y. O., & Hosseini, M. (2005). Towards Resilient Communities in Developing Countries Through Education of Children for Disaster Preparedness. *Int. J. Emergency Management*, *2*.

Malmi, E., Takala, P., Toivonen, H., Raiko, T., & Gionis, A. (2015). DopeLearning: A Computational Approach to Rap Lyrics Generation. https://doi.org/10.1145/2939672.2939679

Mcclain O., L. (2015). Music as Education, Voice, Memory, and Healing: Community Views on the Roles of Music in Conflict Transformation in Northern Uganda. https://doi.org/10.2979/africonfpeacrevi.5.1.41

Matthew, Jr., & Melvin E. (2011). Duck and Cover: Civil Defence Images in Film and Television from the Cold War to 9/11. North Carolina: McFarland & Company. Inc.:11-12.

Nonaka I., & Takeuchi H. 1995. The Knowledge-Creating Company. Oxford University Press: 57-71

Syafwina. (2014). Recognizing Indigenous Knowledge for Disaster Management: Smong, Early Warning System from Simeulue Island, Aceh. https://doi.org/10.1016/j.proenv.2014.03.070

The Guardian. (2018). Japan Earthquake: 100,000 Children Displaced, Says Charity | World news | The Guardian. Retrieved October 27, 2018, from https://www.theguardian.com/world/2011/mar/15/japan-earthquake-

children-displaced-charity

The Straits Times. (2019). 15 Students In Pasir Gudang Suffer Breathing Difficulties Possibly Due to Toxic Gas, Retrieved July 22, 2019, from https://www.straitstimes.com/asia/se-asia/15-students-in-pasir-gudang-suffer-breathing-difficulties-possibly-due-to-toxic-gas

Trstenjak, B., Mikac, S., & Donko, D. (2014). KNN with TF-IDF Based Framework for Text Categorization. *Procedia Engineering*, *69*, 1356–1364. https://doi.org/10.1016/j.proeng.2014.03.129

Tsai, M.-H., Chang, Y.-L., Kao, C., & Kang, S.-C. (2011). The Effectiveness of a Flood Protection Computer Game for Disaster Education. https://doi.org/10.1186/s40327-015-0021-7

Tuswadi, & Hayashi, T. (2014). Disaster Prevention Education in Merapi Volcano Area Primary Schools: Focusing on Students' Perception and Teachers' Performance. https://doi.org/10.1016/j.proenv.2014.03.080

Walton, P. (2016). Using Singing and Movement to Teach Pre-reading Skills and Word Reading to Kindergarten Children: An Exploratory Study. *Language and Literacy*, *16*(3), 54. https://doi.org/10.20360/g2k88j