International Journal of Engineering & Technology, 7 (4) (2018) 2192-2196

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET
doi: 10.14419/ijet.v7i4.15149
Research paper

Integrated-system to minimizing cyber harassment in kingdom of Saudi Arabia (KSA)

Fahad Abdullah Moafa [1] *, kamsuriah Ahmad [1], Waleed Mugahed Al-Rahmi [2], Noraffandy Yahaya [2], Yusri Bin Kamin [2], Mahdi Alamri [3]

[1] Faculty of Information Science and Technology, University Kebangsaan Malaysia, Malaysia.
[2] Faculty of Education, Universiti Teknologi Malaysia, 81310, UTM Skudai, Johor, Malaysia
[3] Faculty of Education, education technology department, king Faisal university, Alahssa, Saudi Arabia
**Corresponding author E-mail: fah171393@hotmail.com, waleed.alrahmi@yahoo.com*

Abstract

The proposed system framework consists two main databases: Lexicon dictionary and Summarized previous cases, by depending on Senti-ment analysis and N-Gram algorithms to match the terms and documents. In the first branch, the judge opens the cyber case and therefore the system will highlight the technical terms automatically. Furthermore, the technical terms matched with Lexicon dictionary will be high-lighted. After that, the judge opens the highlighted terms (as links), and description page will be appeared. The description page contains details about the technical terms (definitions, explanations, examples, etc). On the other side, the second branch aims to retrieve the related legal cases (from the database) judged by courts in UK and KSA. The related cases are the most closed cases to the current legal case by inserting keywords based on the current case. The judge benefits from these cases through the judgment issued to give the fair judgment. N-gram algorithm is used to find the related cases because it has smart approach to expect the most closed document and texts. The system provides the judge with laws used in issuing the judgment in KSA and UK courts.

Keywords: *Integrated System; Cyber Harassment; Students' in (KSA*

1. Introduction

According to [1] focused on detecting cyberbullying on the social media networks. The researchers highlighted that the social media networks are suitable cyber space for the harassers. Detecting cyberbullying according to the researchers includes two features: NUM and NORM, where NUM is a count and NORM is a normalization of the bad words. They employed replication of positive examples up to ten times and they reported the accuracy on the range of classifiers. However, on the first stages they used semantic features to detect harassment. Also, they illustrated the appearance of pronouns and they used types of features sets: all second person pronouns such as 'you', 'yourself, other remaining pronouns such as 'he', 'she', and foul words such as 'stupid'. The study of [2] highlighted the term of cyberbullying and indicated the technical aspects of this crime. Also, it aimed to introduce an approach for detecting and combating cyber harassment and cyberbullying.

By using text mining, the identification of sexual predators and other crimes like vandalisms, spam internet abuse and cyber terrorism becomes feasible online [3]. Furthermore, the proposed approach includes four main steps: selecting different profiles, collecting profiles data and tweets, features, and supervised learning. In contrast, [4-6] students and researchers have positive attitudes and intentions to use social media for educational purposes.

The problem of detecting the cyber harassment is located in using the foul words as indicator of friendship and close relationship, specialty by teenagers. Therefore, proofing cyber harassment, in addition to finding the victim's response is so difficult. The system suggestion in [7] is basically relies in mining paradigms, like discovered online sexual harasser, which basically depends on collecting data through the use of annotations retrieved through platforms such as MTurk or Crowd flower in crowd Sourcing. As a first step, the same user gets identified in the different networks, their behavior are observed and their reaction is also assessed after a case of harassment if or not the harassment is related to case of bullying or not, figure [8, 9] shows the conceptual model of the proposed approach. The employee support vector machine (SVM) collects the Auxiliary information and the effect of gender-specific language structures on the process of detecting cyberbullying in social networks is carefully studied. This is done under the assumption that the overall detection accuracy could be enhanced as the gender-specific language structures are collected. This contrasts with previous studies [10], [11] where social media is used for engagement among students. See figure 1.

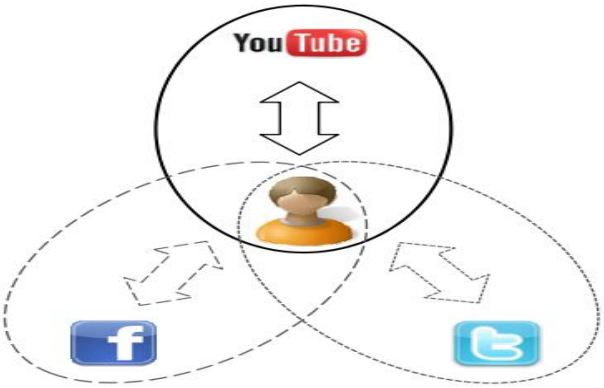

Fig. 1: The Conceptual Model [7].

2. Technical part

The current study only targeted the usage of pronoun and vocabulary usage differences. This is done in order to prove the assumption in which the accuracy of classification in terms of harassing contents can be enhanced through the use of gender-specific features. The current study is also characterized by the use of foul words in 100,000 posts that were randomly selected the dataset and compared the foul words used in a frequent basis by each gender. According to the results of Wilcoxon signed rank test, it is found that there is a significant difference between males and females in terms of the frequency in using the foul words in their posts fig 2 illustrates Top ten frequently used foul words by female (circle) versus male (square).

Fig. 2: Top Ten Frequently Used Foul Words According to Gender [7].

On the other hand the researchers in [12] pore in studying cyberbullying and cyber harassment. They used analyzing language to detect cyberbullying appeared in the texts at social networks or SMS. The researchers aimed to find a ways for detecting instances of cyberbullying by generating a model of language on the light of the text in online posts. Also, they used machine learning algorithms for detecting cyberharassment. In this study, the researchers introduced specific terms that are indicative of cyber harassment.
The study aimed to reduce the repeated characters within the word such as looooooooooove will become love. Also, the method of Lemur was used in order to index the data and produce the term by document matrix. This study can contribute to our work by providing the terms that are indicative of cyber harassment. On the other hand, the researchers in [13] declared cyber harassment and cyber stalking. Also, they indicated the technical aspects of cyber stalking and cyber harassment. Since the new characteristics of the digital generation is it's excessive consumption which inhibits capital accumulation and market balance, which contributes significantly in new enhanced applications, programs, and games that may contain malicious programs. Thus, it is serious to raise the security issues when building the applications. On the other side, studying cyber harassment is considered significant argument as it was mentioned by [14-15].

3. Framework of the proposed novel integrated system

As shown in figure 6, the proposed system is smart because it summarizes the legal case with the judgment and laws that were depended, and then stores the document in the database of "Summarized cases", in order to make system more experts, and so on. See Figure 3.

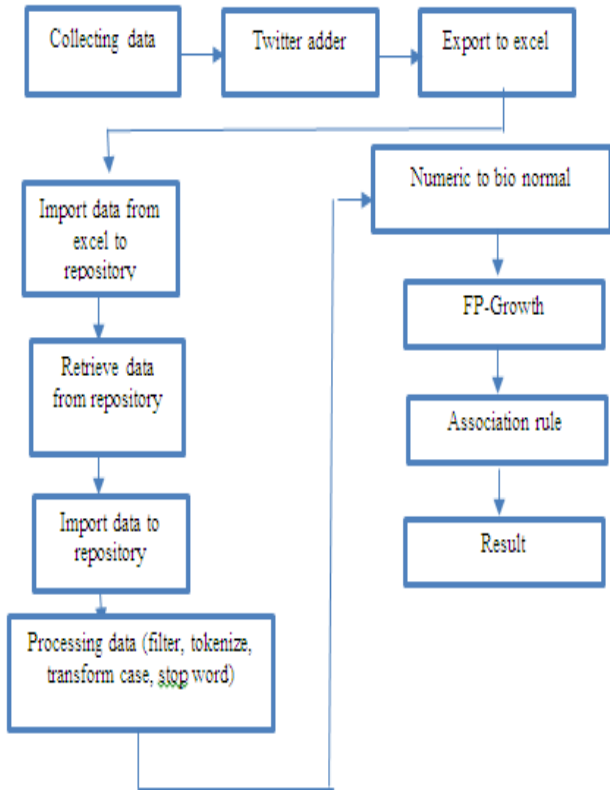

Fig. 3: The Process Analyzing Indonesian Bullying Words From Indonesian Twitter [15].

Furthermore, the study talked about Indonesian cyber bullying patterns by introducing an approach of mining bullying pattern on Indonesian Twitter post as shown in figure 3. Moreover, the procedure consisted of two main steps: utilizing FP-Growth in terms to conduct analysis tweets of retrieving frequent patterns item sets; and analyzing the robust relationship between a pair of words using association rules mining. The study divided FP-Tree into the following steps:
First, finding the frequent item sets by counting and scanning the data. Second, renounce the items with frequency in a decreasing order. Third, establishing FP- Tree and scanning every transaction from the data set. This happens by creating a new path in case the transaction form and set the counter for each node to be 1, and adding the common item sets node counters if the transaction join in common prefix item sets then and establish new node.

4. Mathematical background

4.1. N-gram

N-Grams are known as an algorithm for anticipating that uses probabilistic methods in anticipating next word after observing N-1 words. Thus, there is a strong relation between computing the probability of a sequence of words and that of the probability of the next word.
Word sequences

$$w_1^n = w_1 \; w_n \tag{1}$$

Chain rule of probability

$$P(w_1^n) = P(w_1)P(w_2 \,|w_1)P\left((w_3 \,|w_1^2\right) \ldots P(w_n|w_1^{n-1}) = \prod_{k=1}^{n} P(w_k \,|w_1^{k-1}) \tag{2}$$

Bigram approximation

$$P(w_1^n) = \prod_{k=1}^{n} P(w_k|w_{k-1}) \tag{3}$$

N-gram approximation

$$P(w_1^n) = \prod_{k=1}^{n} P(w_k | w_{k-N+1}^{k-1}) \quad (4)$$

The rule illustrated above highlights the link between computing the conditional probability of a word based on previous words and the computing of the joint probability of a sequence. The fourth equation above illustrates that joint probability of an entire sequence of words can be estimated through multiplying a number of conditional probabilities.

4.2. Sentiment analysis

The system will divide the text into individual word to prepare the text for comparison operation. System will consider special algorithm for making division operation in order to make each word individual one. The following pseudo code shows the division operation.

Input: Text T
Variable: W word
Output: list array of words
Read input text
W ← gets word from text
Int arr[ ] as array
For each w to end of text If read character is "NULL" "which is space" or special character"* /? > etc." Then Consider it is the end of the word Else continue
End For

Most of the discussed articles pinpoint the automatic deriving of word Sentiment Orientation SO out from existing linguistic data. The issue of the quality and how these derived SO values can be utilized in text classification tasks has received little attention by researchers. One of the expectations is that a two-word bigrams approach, which can be extracted according to their part of speech (i.e., adjective/noun pairs, adverb/verb pairs, etc.), can be used instead of unigram (single word) approach.

Through the calculation of the Pointwise Mutual Information (PMI) of these bigrams, their SO values can be derived. Below is the definition of this term[9]:

PMI (word1, word2) = log2 (p (word1 & word2)/ p (word1) p (word2))

That is, "the PMI of two words is equal to the base-2 log of the probability of the two words appearing together, divided by the product of the independent probabilities of the words; as such, the PMI of two words that appear independently of one another would be close to zero

(since p (word1 & word2) = p (word1) p (word2))". (5)

The researchers mentioned that an online grooming attack is an adult that approaches the children online with the goal of performing sexual activities and contains sexual conversation. The study interested with analyzing images and videos on the social media networks particularly Facebook. The first step is face recognition application using Haar like and Viola-Jones. The next step is computing facial and skin regions determined by skin-detection module. This method depended on percentage of finding skin to non-skin in an image a classification in pornographic or non-pornographic. This study is effective in introducing detecting pornographic based on Haar-like and Viola-Jones algorithms. On the other side, [16] highlighted that biased and harmful contents appeared to negatives influence teenagers if compared with adults.So, the researchers introduced an approach to protect adolescents from cybercrimes. The proposed system contained preprocessing and two major components: sentence offensiveness prediction and user offensiveness estimation. User's conversations are chunked into posts in the stage of preprocessing. Also, they derived each sentence's offensiveness from two lexical features for representing words' offensiveness in a sentence. Lexicon features to represent words' offensiveness in a sentence. For each word, they compute intensify value d using the following equation:

$$\begin{cases} b1 \text{ if cj is a user identifier} \\ b2 \text{ if cj is an offensive word} \\ 1 \text{ otherwise} \end{cases} \quad (6)$$

Where b1>b2> 1, indicates that the level of aggressiveness in higher in the words describing users more than they are in the words that describe other aggressive words.

Table 1: The Proposed Rules

Rules	Meanings	Examples	Dependency Types
Descriptive modifiers and components: (noun, verb, adj) ← (adj, adv, noun)	B used to modify A	You f**king;	• Obbrev (abbreviation) • acomp (adjectival complement) • amod
Object: : B(noun, verb) ← A(noun, verb)	A is B's direct or indirect object.	F*** yourself; Shut the f** up;	Dobj (direct object) Iobj (indirect object)
Subject: A (noun) ← B(noun, verb)	A is a B's subject or passive subject.	You f** …; You are f**	Nsubj (nominal subject) Nsubjpass(passive nominal subject)
Close phrase, coordinating conjunction: A and B A, B.	A and B or two Bs are close to each other in a sentence.	• F** and stupid;	• Conj(conjunct) • Parataxis(from Greek for place side by side)
Possession modifier: A(noun)→B(noun)	A is possessive determiner of B	You f***.	Poss (holds between the user and possessive determiner)

Table 1 shows the syntactical intensifier detection rules which generated by the researchers to compute the level of offensive. This study is considered effective to our work because it contributes with rules of finding offensive words. Thus, the proposed method can be used to detect cyber harassment.

On the other hand, the study of [17] introduced an approach called Predator and Prey Alert (PAPA) for combating cyberstalking and cyber harassment. Firstly, the researchers considered that a high bandwidth connection is linking the victim to the internet. Also, the victim uses instant messaging or other mean in order to communicate with other individual. Furthermore, they supposed that the stalker harasses the victim in chat rooms. The victim checks email and the threat is contained in one or more emails. The researchers claimed that PAPA is responsible for capturing all videos information and capturing other meta-data. PAPA components include session recorder, victim module, agent module which is a software module that makes it possible for the CUI to be monitored by an investigator. It also allows a dispatcher to authenticate connections between PAPA components through monitoring the state of the session. On the other side, [18] introduced an approach depended on supervised machine learning for detecting cyberharassment and cyberbullying. The proposed approach includes the selecting random Twitter profiles, where the researchers selected 19 different profiles. The idea of approach was to gather different profiles with social relations in social networks and in real life. WEKA platform was used because of the classification algorithms it contains. Random Forest, K-Nearest, and Sequential mining

were used. For optimizing the results, the researchers filtered the tweets into with stop words in Spanish language.

5. Implementations and conclusions

The researchers in [19] introduced a novel approach for determining the author of malicious emails. Also, they aimed at reaching a proof that can support the conclusion on authorship. Author Miner is a new data mining methodology that was proposed by the researchers. Its objective was to check the authorship of a malicious e-mail μ from a group of suspects $\{s1, \dots, sn\}$ on the light of the extracted features of their previously written e-mails $\{E1, \dots, En\}$. They used three main phases: mining frequent patterns, filtering common frequent patterns, and identifying author. In the first phase the patterns $FP\ (Ei)$ are extracted from each collection of e-mails Ei written by suspect Si.

The second phase represented the writing patterns of suspect Si that has been captured by $FP(Ei)$. It is possible that this $FP(Ei)$ includes frequent patterns that are similar or mutual to other suspects. In the third phase, the malicious e-mails µ are compared with each write-print $WP(Ei) \in \{WP(E1), .., WP(En)\}$ and identify the most similar write-print that matches µ. Furthermore, identifying author of the malicious e-mail µ is done by comparing µ with each write-print $WP(Ei)$ and identifying the most similar write-print to µ. Equation 1 shows the score function that was used to quantify the similarity between the malicious e-mail µ and a write-print $WP(Ei)$.

$$score(\mu \approx WP(Ei)) = \frac{\sum_{j=1}^{p} support(MPj|Ei)}{WP(Ei)} \quad (7)$$

Where $MP = \{MP1, ., MPp\}$ is a set of matched patterns between $WP(Ei)$ and the malicious e-mail µ. Hence, the score is a real number within the range of (0, 1), and the highest score is the author of malicious e-mail µ and vice versa. Moreover, this approach benefits our work by providing an algorithm for detecting the writers of malicious e-mails. Thus, this approach can be exploited to discover and detect the harassment words and determine the writers (harasser). The proposed system is robust system because it provides high quality services to the judges. Moreover, the system aids the judge to give justice resolution by providing details about technical terms that appear in the files of cybercrime. Hence, it is possible to reduce time consumed by the judge in studying the cyber-crime file and the judge will be satisfied because he will give justice resolution. The propose system is based on two main algorithms; the first one is N- gram algorism. This algorithm is characterized from other by encoding not just keywords, but also word ordering, automatically, models are not biased by hand coded lists of words, which exactly what we need in system and what judges need [20], [21]. The second algorithm is Sentiment Analysis which depends on lexicon dictionary and has many benefits such as: creating questions based on the goals, working assess all comments in the identified channels and giving real time results, on other hand, the lexicon dictionary contains all technical words judge may needs with sample explanation. In conclusion, the esigning, implementing, and testing a novel integrated system for the authorities in Saudi Arabia (KSA), where this system describes a novel methodology for developing reporting and evidences architecture, aims to support evidence keeping and instant reporting of incidents to the authorities systematically.

On the other hand, the proposed system is smart enough because it makes auto experience through storing the current cybercrime in the database of related cases in order to benefit from this case later on. Each case is stored with name and resolution issued in specific date. For future work apply and test an integrated system to maintain harassment reporting and evidence sharing with the authorities. The system is expected to be following client-server architecture with gateways such as mobile phones interconnected through the web.

Acknowledgements

We would like to thank Faculty of Information Science and Technology, Universiti Kebangsaan Malaysia by giving the authors an opportunity to conduct this research. This research is funded by Universiti Kebangsaan Malaysia under Exploratory Research Grant Scheme ERGS/1/2013/ICT07/UKM/02/1. Also, we would like to thank the Research Management Centre (RMC) at Universiti Teknologi Malaysia (UTM) for funding this project under grant number PY/2017/00760: Q.J130000.2531.16H59.

References

[1] Author,"Title of the Paper", *Journal name*, Vol.X, No.X, (200X), Nahar, V., Li, X., Pang, C., (2013) An Effective Approach for Cyberbullying Detection, Vol. 3 ISSN.5, PP. 238-247.

[2] Laorden, C., Galán-García, P., Santos, I., Sanz, B., Hidalgo, J. M. G., & Bringas, P. G. (2013, January). Negobot: A conversational agent based on game theory for the detection of paedophile behaviour. In International Joint Conference CISIS'12-ICEUTE´ 12-SOCO´ 12 Special Sessions (pp. 261-270). Springer Berlin Heidelberg.

[3] Laorden, C., Sanz, B., Alvarez, G., Bringas, P.G. (2010) A threat model approach to threats and vulnerabilities in on-line social networks. In: Computational Intelligence in Security for Information Systems. Volume 85 of Advances in Intelligent and Soft Computing.

[4] Al-rahmi, W. M., Othman, M. S., & Yusuf, L. M. (2015). Social media for collaborative learning and engagement: Adoption framework in higher education institutions in Malaysia. Mediterranean Journal of Social Sciences, 6 (3 S1), 246-252. https://doi.org/10.5901/mjss.2015.v6n3s1p246.

[5] Al-Rahmi W. M, & Zeki, A. M (2017). A model of using social media for collaborative learning to enhance learners' performance on learning. Journal of King Saud University-Computer and Information Sciences, 29(4): 526-535 https://doi.org/10.1016/j.jksuci.2016.09.002.

[6] Al-rahmi, W. M., Othman, M. S., & Yusuf, L. M. (2015). The effectiveness of using e learning in Malaysian higher education: A case study Universiti Teknologi Malaysia. Mediterranean Journal of Social Sciences, 6(5), 625- 637. https://doi.org/10.5901/mjss.2015.v6n5s2p625.

[7] Dadvar, M. and De Jong, F. (2012) Cyberbullying detection: a step toward a safer Internet yard. In Proceedings of the 21st international conference companion on World Wide Web.121-126.

[8] Moafa, F. A., Ahmad, K., Al-Rahmi, W. M., Alias, N., & Obaid, M. A. M. (2018). Factors for Minimizing Cyber Harassment among University Students: Case Study in Kingdom of Saudi Arabia (KSA). Journal of Theoretical & Applied Information Technology, 96(6).

[9] Turney, P. (2002) Thumbs up or Thumbs down? Semantic Orientation Applied to Unsupervised Classification of Reviews. In Proceedings of the 40th Annual Meeting on Association for Computational Linguistics, pages 417_424. Association for Computational Linguistics.

[10] Al-rahmi, W. M., Othman, M. S., & Yusuf, L. M. (2015). Exploring the factors that affect student satisfaction through using e learning in Malaysian higher education institutions. Mediterranean Journal of Social Sciences, 6(4), 299. https://doi.org/10.5901/mjss.2015.v6n4s1p299.

[11] Al-rahmi, W. M., Othman, M. S., & Yusuf, L. M. (2015). The effect of social media on researchers' academic performance through collaborative learning in Malaysian higher education. Mediterranean Journal of Social Sciences, 6(4), 193-203. https://doi.org/10.5901/mjss.2015.v6n4s1p193.

[12] Kontostathis, A., Reynolds, K., Garron, A. and Edwards, L. (2013) Detecting cyberbullying: query terms and techniques. In Proceedings of the fifth Annual ACM Web Science Conference. 195-204. https://doi.org/10.1145/2464464.2464499.

[13] Kim, T. Y., Jung, M. A., Kim, E. and Heo, E. (2014) the Faster Accelerating Growth of the Knowledge-Based Society. In Economic Growth. 193-235. Springer Berlin Heidelberg.

[14] Moafa, F. A., Ahmad, K., Al-Rahmi, W. M., Yahaya, N., Kamin, Y. B., & Alamri, M. M. (2018). Cyber Harassment Prevention through User Behavior Analysis Online In Kingdom of Saudi Arabia (KSA). Journal of Theoretical & Applied Information Technology, 96(6).

[15] Margono, H., Yi, X. and Raikundalia, G. (2014) Mining Indonesian Cyber Bullying Patterns in Social Networks.
[16] Church, K. and Hanks, P., (2000) Word association norms, mutal information and lexicography. Paper presented at 27th Annual Conference of the ACL, New Brunswick, NJ, [online]: 1 April 2013, available at: http://acl.ldc.upenn.edu/J/J90/J90-1003.pdf.
[17] Chen, Y., Zhou, Y., Xu, H., (2012) Detecting Offensive Language in Social Media to Protect Adolescent Online Safety. International Conference on Social Computing, IEEE Journal.
[18] Aggarwal, S., Burmester, M., Henry, P., Kermes, L., (2005) Anti-Cyberstalking: The Predator and Prey Alert (PAPA) System. Journal of IEEE.
[19] Garcia, P., Gaviria, J., (2011) Supervised Machine Learning for the Detection of Troll Profiles in Twitter Social Network: Application to a Real Case of Cyberbullying.
[20] Jaishankar, K. (Ed.). (2011) Cyber criminology: exploring internet crimes and criminal behavior. CRC Press. https://doi.org/10.1201/b10718.
[21] Chambers, N., Tetreault, J. and Allen, J. (2004) Approaches for automatically tagging affect. In Proceedings of the AAAI Spring Symposium on Exploring Attitude and Affect in Text: Theories and Applications.