

Examining Cyber-Bullying Behavior on Social Networking Sites in general

Bashayr Alqabasani
College of Engineering,
EFFAT University
An Nazlah Al Yamaniyyah,
Jeddah, 22332, Saudi Arabia
baalqabasani@effatuniversity.
edu.sa

Maha Alghamdi
College of Engineering,
EFFAT University
An Nazlah Al Yamaniyyah,
Jeddah, 22332, Saudi Arabia
mbalghamdi@effatuniversity.
edu.sa

Maram Bashawyah
College of Engineering,
EFFAT University
An Nazlah Al Yamaniyyah,
Jeddah, 22332, Saudi Arabia
mabashawyah@effatuniversity.
edu.sa

Adel M. Ilahi
College of Engineering,
EFFAT University
An Nazlah Al Yamaniyyah, Jeddah, 22332, Saudi
Arabia
aiilahi@effatuniversity.edu.sa

Joud Abdulfattah
College of Engineering,
EFFAT University
An Nazlah Al Yamaniyyah, Jeddah, 22332, Saudi
Arabia
jaabdulfattah@effatuniversity.
edu.sa

Article Info

Volume 81

Page Number: 4064 - 4068

Publication Issue:

November-December 2019

Article History

Article Received: 5 March 2019

Revised: 18 May 2019

Accepted: 24 September 2019

Publication: 19 December 2019

Abstract:

Bullying is defined as a form of violence and abuse directed at a person or group of persons. Cyber-bullying is the frequent and violent abuse of people through the Internet. The behavior is prevalent in social networking sites and aims at harming and threatening the target person/victim. Cyber bullying can reveal personal information about the victim (real name, home, work or school address) on websites/forums, or use impersonation (including fake sites, comments, accounts etc.). The form and extent of cyber-bullying is unknown in social networking sites. Therefore, this study aims at uncovering this information by study the cyber-bullying behavior online and addresses its prevalence in the Arab world. Web Crawler is used to collect all the data relate to bullying on social media and the result demonstrated a high connection between the utilized watchwords and the constructed corpus.

Keywords:- Cyber-Bullying; social media; social networking; web crawler

1. INTRODUCTION

Cyber-bullying is the electronic communication that people use it to bully to other people by using email or social media to send bad message or comments such as Twitter, Facebook, etc. There are many kinds of bullying which is physical assault, teasing, or making threats. Cyber-bullying has become widespread through

junior high schools. Moreover, Cyber-victimization is associated with negative psychological health outcomes [1]. The study was about cyber-bullying among Saudi's higher-education students. Around 27% of the students reported that they have committed cyber-bullying, 57% of the students observed at least one student being cyber-bullied [2].

Cyber-bullying has been recognized as an imperative issue among youth. A few illustrations specified of definitional issues, redundancy and power unevenness, sorts of digital tormenting, age and sexual orientation contrasts, cover with conventional harassing and the arrangement of occasions, the contrasts between digital harassing and customary harassing, thought processes in and effect of digital exploitation, adapting systems, and counteractive action/intercession conceivable outcomes [3].

Digital cyber-bullying is bullying via social media, email, and text messaging and other electronic means. Like traditional, in-person bullying, cyber-bullying is an act of aggression characterized by a power imbalance. So, many young people today are experiencing acts of cyber-bullying. Kowalski et al (2014) study is based on two things (1) to provide of the surviving exploration on cyber-bullying among youth, including an investigate the commonness and precursors of this conduct and related results; (2) to synthesize the connections among cyber bullying, cyber victimization, and significant behavioral and mental factors with meta-analytic [4].

Technological bullying or cyber-bullying, has allowed the problem to expand, and harder to define [5]. Anonymity has been viewed as one of the different between cyber-bullying and traditional bullying. The anonymity is an important risk factor for later cyber-bullying behaviour [6]. Anonymity is the primary contributor to cyber bullying and directly affects online behaviour as its easier approach and difficult to detect [7]. Therefore, the detection of the cyber-bullying is important.

In web corpus construction, Crawling is considered as an essential process due to its expensive bandwidth requirements that result in an increase in storage capacity. Roland Schafer et al. (2014) have prepared three examinations to enhance the ratio between corpus size and crawling effort for web corpus construction. The results are based on certain scores: quality of general content on the page contains a link while the other one is related to the probability of a block containing a link with boilerplate detection [8]. It has a high potential for streamlining web corpus crawling through the prioritization of collecting URLs in a crawler framework [8].

Web crawling framework to focus on sentiment is proposed by organize found URLs in light of their anticipated assessment scores and managed to discovery the sentimental Web content [9]. The web crawler works in two steps, first will fetch the relevant sites, and then will retrieve the relevant sites through deep search by in-site exploring [10]. Therefore, this study uncovering the cyber-bullying information through Web Crawler in order to gather the important data relate to bullying on social media.

2. Methodology

This study considered Twitter with the purpose to assess the adequacy of the corpus for online networking. In particular, the

issue to distinguish the keywords particular to cyberbullying is limited. Therefore several assumptions are defined as following:

- a. Different kinds of cyberbullying was separated (e.g., the words are totally different from KSA and US and its meaningful.)
- b. This study focus around corpora in the Arabic and English dialect and consequently our outcomes will just apply to English talking group (However, we don't have any kind of effect amongst local and non-local speakers).
- c. This study fabricates our corpora utilizing a given arrangement of hash tags that particularly discuss cyberbullying.
- d. Cyberbullying may affect the psychological aspect and disturb the mood of the person.

2.1 Data Sources and Data Gathering

The information related to cyberbullying and awareness in society is considered in this study. The user's accreditations are confidential and anonymize by giving imaginary identifiers. Besides that, many people who are cyberbullying don't utilize their genuine name or put fundamental data about themselves. The assembled data is originates from the accompanying sources, which are:

- a. Standard Corpus – Comprised of essays on cyberbullying and give a premise to affirm the suspicion
- b. Set of known hashtags that discussion about cyberbullying.
- c. Tweets composed by specific user who has partaken in the hashtags.

2.2 Preprocessing and Processing Data

After gathering the tweets, the following steps are performed.

- a. Removing the stop words: This implies evacuating words, for example, "for", "to" and so forth.
- b. Vectorize the rest of the words: In basic terms, utilize the well-known term recurrence calculation to remove the catchphrases and utilize the tf-idf calculation executed in python. For each word, ascertain the recurrence of the considerable number of words
- c. The most recent two months of tweets for ten arbitrary clients is accumulated (500 tweets if two months information isn't accessible). The decision of the quantity of months is self-assertive and can be changed in the program effectively. If no ton of tweets on Twitter is accumulated in a week, so it takes more time to get the normal outcomes.
- d. Once the information is accumulated, the user ids are anonymized for detailing purposes to keep up the protection of the clients.

e. The information is stored in a NoSql database - MongoDB.

2.3 Evaluation

The catchphrases in those hashtags are further analysed. In particular, the hashtags produced two arrangements of catchphrases to be specific

- 1) Standard lexicon words that the clients wrote and
- 2) Out of Vocabulary words (OOV) that can't be found in a word reference.

The recurrence of comparative lexicon words in the hashtags against those in the standard corpora gathered from Depression writing. The hidden supposition was that if the recurrence of models word in hashtag bears a high connection then by associativity the OOV words may speak to the pervasive utilization in online networking. The adequacy of base corpus is developed and utilized the corpus against the irregular ten clients that picked who took an interest in the base arrangement of hashtags. During monitoring the words that every specific client contributed toward the corpus, those words from the corpus are expelled.

2.4 Implementation of web crawler

VirtualBox equipped with web crawler is used in this study. Terminal Emulator is used for send command that to be implemented. Terminal imitating is the capacity to influence an offered PC to seem like a real terminal or customer computer arranged to a server or centralized computer. This allows obtains the data or projects on the server or centralized computers, which are normally just accessible to the terminal being copied

The cd command shown in Figure 1, otherwise called chdir (change directory), is a summon line OS shell charge used to change the present working registry in working systems, for example, Unix, DOS, OS/2, Amiga OS (where if an uncovered way is given, cd is suggested), Windows, and Linux. It is additionally accessible for use in shell contents and clump records. The system call that influences the order in most working systems is chdir that is characterized by POSIX. The code used is (cd /home/maram/Downloads/) which enable to see and open any file in the Downloads file.

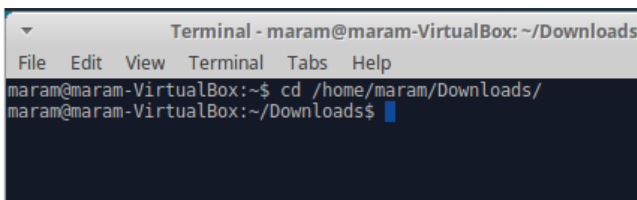


Figure 1. cd comment in Terminal Emulator

In computing, ls are a command to the list of documents in UNIX and UNIX-like of the operating systems. LS are indicated by POSIX and the Single UNIX Specification. At the point when

summoned with no contentions, LS list the documents in the present working directory. Use (ls) to show the files in the Downloads folder. And here it show that there is file which called crawl.py. But it show the crawl.py file is still not open it. This file is the code that can use it to find and search of the tweets from the Twitter.

Although su can be utilized to change the responsibility for session to any client, it is most ordinarily utilized to change the proprietorship from a conventional client to the root (i.e., managerial) client, along these lines giving access to all parts of and all commands on the computer or system. Thus, it is regularly alluded to (albeit to some degree erroneously) as the superuser command. It is likewise now and then called the switch user command. Wrote (su -) to allow enter as root and it require entering the password. After enter of the root and it will allow using it.

Install python by using code (sudo apt-get install python-pip). Pip is packages that will management system and it will be used to install and manage software or programming the packages written in Python. Numerous packages can be found in the real default source for the packages and their conditions — Python Package Index (PyPI). Pip is also called a recursive acronym that can be remaining for either of "Pip Installs Packages" or "Pip Installs Python". On the other hand, pip can remains for "favored installer program". Then wrote (pip install twitter) to installing the twitter program and it shows that was installed successfully as shown in Figure 2. It can help to search and find the tweets.

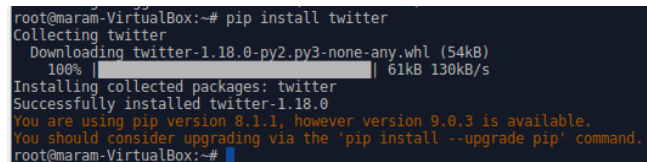


Figure 2. Installation of twitter

Python is a deciphered abnormal state programming language for broadly useful programming. It gives builds that empower clear programming on both little and huge scales. Python includes a dynamic sort framework and programmed memory administration. It underpins different programming ideal models, including object-situated, basic, useful and procedural, and has a large and exhaustive standard library. The code that wrote it here is (python crawl.py) to open the file which called crawl.py that was saved in downloads folder and then can use it after its open. Code (ls) its show that the file which called crawl.py is open.

More command can likewise be utilized for perusing the output of different commands/applications in the console. In this picture, wrote code (more outputtweet_20180403_072353.txt) to show all tweets that the people entire it in the Twitter. And the picture shows some of the tweets not all of them. The tweets are depending on the hashtag that you want to look about it and these

tweets are about #التتمر or #cyberbullying. The tweets in Arabic and English hashtag form the tweets are compare between them.

It is accessible on UNIX and Unix-like systems, DOS, OS/2, and Microsoft Windows. Projects of this sort are called pagers. More is an exceptionally fundamental pager, initially permitting just forward route through a document, however more up to date usage do take into account constrained in reverse development. This study wrote the code (more crawl.py) to show all the information of Twitter like the id, hashtag, and others.

3. Result and discussion

3.1 Keyword Identification for Standard Corpus

The keywords that were used in both the standard corpus and the social media were identified. Utilizing the tf-idf algorithm, essays on Cyber-bullying, Bullying, and Psychological Cases were recognized and the top twenty words are shown in Table 1.

Table 1. Survey Questionnaire Responses

Keywords	
Attack	Trolls
Teen	Comments
Protecting	Cell
Humiliation	Threats
Impact	Isolation
Mental	Cope
Cyberbullying	Help
Health	Community
Behaviours	Harassment
Rumors	Cyberstalking

The considered standard corpus is geared towards general audience and hence the keywords identified were in line with the general audience comprehension.

Certain words identified were not typical when it comes to the discussion regarding cyberbullying such as the word “Rumors” displays in Table 1. However, when taken into context of cyberbullying, it becomes relevant as it is used to describe other relevant symptoms such as stop untrue or harmful messages from spreading before it goes viral because it will harm someone. For example,

- “humiliation”: Cyberbullying occur when someone humiliation others “Comments”: Harassing others by write bad comments
- “Impact”: Bullying has negative impact on people
- “Rumors”: Spread Rumors the false and wrong

3.2 Standard Corpus versus Social Media

Once established the base vector of keywords, the hashtags is considered as base set for Twitter namely #cyberbullying, #bullying and #feelingdown. The #feelingdown hashtag might not necessarily be indicative of long-term cyberbullying but showed many tweets that could indicate potential prevalence of long-term cyberbullying.

Next, sample size of tweets gleaned is defined from the hashtags to 100, 200, 500 and 1000. The following was worth noting:

1. The keywords identified from standard corpus showed a very high correlation to the words identified from the hashtags.
2. For sample sizes of 100, 200, 500, and 1000, the correlation between them was almost 100% in most of words for example in Protecting, Humiliation, Behaviours, Trolls, Cell, Cope, Community, Harassment, and Cyberstalking.
3. Sample size of 100, 200, 500, and 1000 in the others words was unstable and different results of each other e.g., attack. However, as shown in Figure 3, the interesting trend was that for n = 100, the word attack started going up in frequency and was ranked very high in n=100. But in other sample size of 200, 500, and 1000, the word attack started gradually to going down in frequency. This can be explained that as numbers of tweets go up, the users’ choice of words become more narrow and accurate. e.g., “having harm to others in media” versus “attack”. Both terms however could possibly point to the symptom of harm others that can accompany cyberbullying.
4. Despite the non-standard language used on the Internet and Twitter, the keywords identified in both corpora were similar. This result was consistent with the findings where the authors emphasized the following: 1) the narrower the context, better the selection of keywords and 2) despite the presence of OOV words, the underlying text still conforms to the subject at hand.
5. The corpus gleaned from hashtags can serve as a basis to classify tweets from a test data to help detect cyberbullying.

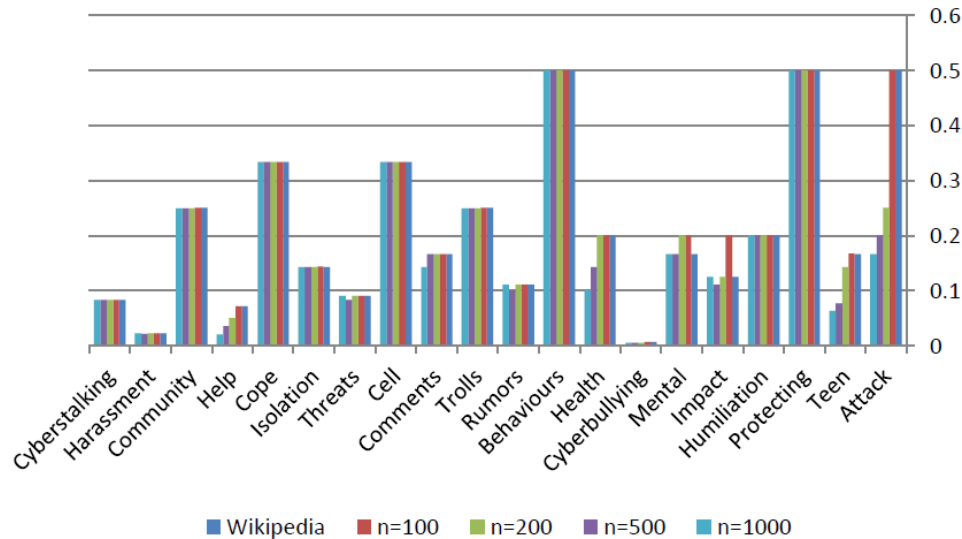


Figure 3. The frequency of keywords detected from various sample size

4. Conclusion

In study demonstrated the setting up of a corpus particular to online networking (social media) with regards to psychological well-being. This corpus will help the specialists to recognize identification for a specific infection. This study utilized cyberbullying as our contextual analysis and showed the high relationship between online networking corpus and standard corpus. The analysis of certain clients in Twitter by dug their tweets for a period, which demonstrated a high connection between the watchwords they utilized and the constructed corpus.

References

1. Safaria, T. 2016. Prevalence and impact of cyberbullying in a sample of Indonesian junior high school students. *TOJET: The Turkish Online Journal of Educational Technology*, 15(1).
2. Al-Zahrani, A. 2015. Cyberbullying among Saudi's Higher-Education Students: Implications for Educators and Policymakers. *World Journal of Education*. 5, 3, 15.
3. Slonje, R., Smith, P. and FriséN, A. 2013. The nature of cyberbullying, and strategies for prevention. *Computers in human behavior*. 29, 1, 26-32.
4. Kowalski, R., Giumetti, G., Schroeder, A. and Lattanner, M. 2014. Bullying in the digital age: A critical review and meta-analysis of cyberbullying research among youth. *Psychological bulletin*. 140, 4, 1073.
5. Donegan, R. 2012. Bullying and cyberbullying: History, statistics, law, prevention and analysis. *The Elon Journal of Undergraduate Research in Communications*. 3, 1, 33-42.
6. Barlett, C., Gentile, D. and Chew, C. 2016. Predicting cyberbullying from anonymity. *Psychology of Popular Media Culture*. 5, 2, 171.
7. Mishna, F., Saini, M. and Solomon, S. 2009. Ongoing and online: Children and youth's perceptions of cyber bullying. *Children and Youth Services Review*. 31, 12, 1222-1228.
8. Schäfer, R., Barbaresi, A. and Bildhauer, F. 2014. Focused web corpus crawling. *In Proceedings of the 9th Web as Corpus workshop (WAC-9)*. 9-15.
9. Vural, A., Cambazoglu, B. and Karagoz, P. 2014. Sentiment-focused web crawling. *ACM Transactions on the Web (TWEB)*. 8, 4, 22.
10. Madhusudan, P. and Poonam, D. 2017. Deep web crawling efficiently using dynamic focused web crawler. *International Research Journal of Engineering and Technology*, 4, 6, 3303-3306.