Cyberbullying Detection
A Survey on Multilingual Techniques

Batoul Haidar
Faculty of Engineering
University of Saint Joseph
Beirut, Lebanon
batoul.haidar@net.usj.edu.lb

Maroun Chamoun
Faculty of Engineering
University of Saint Joseph
Beirut, Lebanon
maroun.chamoun@usj.edu.lb

Fadi Yamout
Computer Science Department
Lebanese International University
Beirut, Lebanon
fadi.yamout@liu.edu.lb

***Abstract*—Cyberbullying is the new form of bullying; executed by electronic media and Internet. Cyberbullying is affecting a lot of children around the world including Arab countries. Awareness for cyberbullying is arising and research is taking place in the fields of cyberbullying detection and mitigation and not just the psychological effects of cyberbullying on the victim. Researches on cyberbullying detection have been done in many languages but none has been done on Arabic language cyberbullying detection until the time of writing this paper. Many techniques are utilized in the area of cyberbullying detection, mainly Machine Learning (ML) and Natural Language Processing (NLP). This paper presents a brief background on cyberbullying and all technologies incorporated under this field; in addition to an extensive survey regarding the techniques and advancements in multilingual cyberbullying detection; and finally proposes a plan of a solution for the problem of Arabic cyberbullying.**

Keywords-Cyberbullying; Machine Learning; Natural Language Processing; Arabic Natural Language Processing

I. INTRODUCTION

Children and teens were subject to physical bullying before the abundance of internet, computers and handheld devices. Nowadays, bullying is performed using cyber technology. Around 50% of the youth of America are suffering from cyberbullying [1]. As for the Arab world: 20.9% of middle-school adolescents report bullying in UAE, 31.9% in Morocco, 33.6% in Lebanon, 39.1% in Oman and 44.2% in Jordan [2].

There is little awareness of cyberbullying in the Arab world. One of the rare reports [3] on cyberbullying states that 60% of Gulf Countries' youth openly admit the presence of cyberbullying amongst their peers. This study also states that only quarter the predators online do bully their victims offline. This means that internet have encouraged three quarters of the predators to bully others, while they wouldn't have considered bullying physically.

Most of the research work dealing with cyberbullying focused on the effects of cyberbullying and dealing with the psychological state of the victims after attacks. Less work was directed towards building up technical methods to detect and stop an ongoing cyberbullying attack, or even to prevent cyberbullying attacks while or before they happen [4].

This paper is a survey of all existing literature in multilingual techniques. It concludes at the end that there is no work done on detecting cyberbullying attacks in Arabic language. Therefore, a system for cyberbullying detection is a possible future system for preventing cyberbullying attacks, by detecting and stopping them. For that system, Natural Language Processing (NLP) will be used to identify and process Arabic words. Then Machine Learning (ML) techniques will be used to classify bullying content.

II. BACKGROUND

A. Cyberbullying

Cyberbullying is defined as the use of Internet, cell phones, video game systems, or other technologies to send or post text or images intended to hurt or embarrass another person or group of people [5]. Some examples of cyberbullying include sending mean or threatening messages, tricking someone into revealing personal or embarrassing information and sending it to others, sending or forwarding private messages to others, sharing explicit pictures with others without consent, starting rumors via text message or online or creating fake online profiles on websites such as Facebook, Myspace, Twitter, etc. to make fun of people [5]. There are several categories of cyberbullying as stated by [6] and [7]:

- Flaming: starting a form of online fight.
- Masquerade: where there is a bully pretending to be someone else, in order to perform malicious intents.
- Denigration: sending or posting gossip to ruin someone's reputation.
- Impersonation: Pretending to be someone else and sending or posting material to get that person in trouble or danger or to damage that person's reputation or friendships.
- Harassment: Repeatedly sending profane and cruel messages.
- Outing: Publishing someone's embarrassing information, images or secrets.
- Trickery: Talking someone into revealing secrets or embarrassing information for the sake of sharing it online.
- Exclusion: Intentionally and cruelly excluding someone from an online group.
- Cyberstalking: Repeated, intense harassment and denigration that includes threats or creates significant fear.

Bullying and cyberbullying leave mental and physical effects on both the bully (*predator*) and the victim. Cyberbullying is more severe than physical bullying due to the fact that it is wider, public, and the victim has nowhere to escape. Victims of cyberbullying reported emotional, concentration, and behavioral issues, as well as trouble getting along with their peers. These victims were more likely to report frequent headaches, recurrent stomach pain, and sleeping difficulties. One out of four students revealed that they felt unsafe at school. They were also more likely to be hyperactive, have conduct problems, abuse alcohol, and smoke cigarettes [8].

B. Machine Learning

Machine Learning (ML) is defined as the ability of a computer to teach itself how to take a decision using availa-

2473-3539/16 $31.00 © 2016 IEEE
DOI 10.1109/EMS.2016.35

ble data and experience [9]. Available Data is known as *Training Data*. Decisions to be taken in ML might be a classification or prediction for new objects or data. The computer classifies a new piece of data by depending on learning algorithms. When the training data is labeled, i.e. classified by human experts, the algorithms depending on these labeled data are called Supervised Learning algorithms [10]. In cyberbullying detection, there could be a corpus of data manually labeled (or classified) by people as either containing harm or not, as described in Section III. When the training data is unlabeled, the algorithms depending on these non-labeled data are called Unsupervised Learning algorithms [10]. They teach themselves how to classify the data based on similarities and differences between data. When both supervised and unsupervised learnings are combined together by using labeled and unlabeled data, to get the most out of both ways, the algorithm is known as Semi-Supervised Learning algorithm [10].

When ML is used to classify a certain object as belonging or not belonging to a certain category, the machine learner is called *Binary Classifier* [11]; for example, in spam email filtering, ML algorithms are used to take decisions against incoming emails and label them as either spam or not spam. A second type is when the task given to the classifier, is to match a certain object against several classes or categories, then it is called *Multi-Class Classifier*. A third type might be predicting a value for an object and is called *Regression*, i.e. predicting a priority level for an incoming email.

There are several ML algorithms available, from which the most frequently used in relevance to the scope of research of this paper will be mentioned.

- Naive Bayes: A probabilistic supervised learning method [12] that mainly calculates the probability of an item belonging to a certain class, depending on metrics obtained from training data. Naïve Bayes algorithm was used in some cyberbullying detection research, such as in [13] and [12]. It was used for sexual predation detection.
- Nearest Neighbor Estimators: A simple estimator [14] that uses distance between data instances, in order to map a certain instance to its closest distance neighbor, thus estimating the class of this instance, this algorithm was used in [15] and [16].
- Support Vector Machine (SVM): Also a supervised algorithm. SVM is a binary classifier that assumes a clear distinction between data samples. It tries to find an optimal hyper plane that maximizes the margin between the classes [10]. SVM was used in many cyberbullying detection systems [17], [18].
- Decision Tree: Decision tree learners use a set of labeled data; thus they are supervised learners. Decision trees classify data using a command and conquer approach. Each leaf of the tree represents a classification class and each arc represents a feature inspected from training data [19]. The C4.5 algorithm is an implementation of decision trees. It was employed in cyberbullying detection by [20] and [15].

ML algorithms are widely incorporated in cyberbullying detection systems as seen in Section III, due to the huge amount of data incorporated in social networking platforms, which makes it hard to be processed by human power, thus comes the need for a machine learner.

C. Natural Language Processing

Natural Language Processing (NLP) is the collection of techniques employed to make computers capable of understanding the natural unprocessed language spoken between humans by extracting grammatical structure and meaning from input [21]. NLP is a branch of Linguistics, Artificial Intelligence and Computer Science [22]. NLP research started with Machine Translation in the late 1940s [23]. Then it spread to other areas of application, such as information retrieval, text summarization, question answering, information extraction, topic modeling, opinion mining [24], optical character recognition, finding words boundary, word sense disambiguation, and speech recognition [22].

According to Chandhana [25], NLP can be divided into three areas; *Acoustic – Phonetic*: where acoustic knowledge studies rhythm and intonation of language; i.e. how to form phonemes, the smallest unit of sounds. Phonemes and phones are aggregated into word sounds. Phonetic knowledge relates sounds to the words we recognize. *Morphological – Syntactic*: Morphology is lexical knowledge which studies sub words (morphemes) that would form a word. Syntactic knowledge studies the structural roles of words or collection of words to form correct sentences. *Semantic - Pragmatic*: Semantic knowledge deals with the meaning of words and sentences, while pragmatic knowledge deals with deriving sentence meanings from the outside world or outside the content of the document [26].

D. Common Sense Reasoning/Sentic Computing

Common sense is the knowledge (usually acquired in early stages of life) concerning social, political, economic and environmental aspects of the society we live in. Common sense usually varies among different cultures and is built from layers of learning experiments we acquire throughout life [27].

Computers do not have common sense reasoning by nature, but there is a research field, known by Sentic Computing [28], that aims towards transforming computers into machines that could feel. This field of research is a multidisciplinary approach to opinion mining and sentiment analysis, which uses common sense reasoning and web semantics in order to inspect the emotions, not just the opinions from certain text. The term 'Sentic' derives from the Latin 'Sentire', the root of words like sentiment and sensations.

E. Performance Measurment

Some evaluation metrics were adapted in Information Retrieval (IR) and then extended to other fields of computer science such as ML. These evaluation metrics are used as measures for the performance of IR and ML systems. The most widely used metrics are *Recall*, *Precision* and *F-Measure*.

- Recall is the proportion of returned documents (or values) which are relevant (or correct) $Rl \cap Rt$ [29] out of all relevant documents returned and not returned [30]. The metric is also known as *Sensitivity* of a system.

$$R=(Rl\cap Rt)/Rl. \quad (1)$$

- Precision is the proportion of returned documents (or values) which are relevant (or correct) $Rl\cap Rt$ [29]. The metric is also known as *Accuracy* of a system.

$$P=(Rl\cap Rt)/Rt. \quad (2)$$

- F-Measure, proposed by van Rijsbergen in 1979, is a weighted harmonic mean of precision and recall. It is a combination between Recall and Precision metrics, which was introduced to overcome the negative correlation between Precision and Recall [31].

$$F\beta=(1+\beta^2)\ PR/(\beta^2P+R). \quad (3)$$

- F1 is a special case of F- measure with $\beta=1$. β Is a parameter to control the balance between Recall and Precision where $0\leq\beta\leq\infty$. When β is set to 0, it implies giving no importance to recall, when β tends to ∞ then no importance is given to precision, and when $\beta = 1$ then Recall and Precision will be given equal importance [32].

$$F1=2PR/(P+R). \quad (4)$$

P*:* Precision, R*:* Recall, Rt*:* Returned documents, and Rl*:* Relevant documents.

III. Previous Work

As stated previously, the research efforts in cyberbullying covered several areas, including the detection of online bullying when it occurs; reporting it to law enforcement agencies, Internet service providers and others for the purpose of prevention and awareness; and identifying predators with their victims. No effort was directed towards detecting cyberbullying in Arabic language. A system will be elaborated to detect cyberbullying attacks in both English and Arabic languages, including Arabish (or Arabizi). Arabizi, sometimes referred to it as the Arabic chat alphabet, is the use of Latin letters in writing Arabic text [33]. This system will use ML techniques for transliteration from Arabic and Arabizi to English characters and for feature selection/extraction and classification, thus the focus was on the previous work done in the areas of cyberbullying detection, ML, feature extraction and cross language transliteration and translation.

A. Cyberbullying Detection

Most of the research done in detecting cyberbullying constituted of either a filtration software or ML techniques. A filtration software has to be employed by social networking platforms, in order to automatically delete or shade profane words [34] [35] [36]; but the filtration method is first limited by its inability for detecting subtle language harassment and second it has to be manually installed [37].

Most work other than filtration methods employs ML techniques, where old corpora of comments or conversations is crawled, whether from Facebook, Twitter, Formspring (a platform similar to Facebook, popular between teens) or even real conversations of sexual attackers [38]. These corpora are used to feed ML algorithms responsible for detecting cyberbullying attacks by building a classification rule from the training set. The obtained classification rule classifies the testing set comments. Such work was done in [20] where the authors crawled data from Formspring and used the Amazons Mechanical Turk [39] for labelling comments. Then they used the learning methods from the Waikato Environment for Knowledge Analysis (WEKA) toolkit [40] to teach and test the model for classifying comments.

The problem of detecting subtle language cyberbullying attacks was tackled by Dinakar et al [4]. They depended on commonsense reasoning in the detection of cyberbullying content. As an example of the commonsense they used: they considered comments of wearing makeup when subjected on Males might indicate the presence of harassment. They built their datasets from both YouTube and Formspring for both training and testing. They used NB, JRip, J48 and SVM for text classifications. For feature sets they used general features, such as a list of unigrams or profane words, tf-idf weighting scheme, Ortony Lexicon for negative affect, Part-of-speech tags for commonly occurring bigrams, and Label Specific Features including frequently used forms of verbal abuse.

The weighting scheme tf-idf is the product of term frequency and the inverse document frequency in the dataset. It involves multiplying term frequency (tf), that represents the number of times a term occurs in a document, by inverse document frequency (idf), which varies inversely with the number of documents to which a word is assigned [41].

Nahar, Li and Pang [37] employed the tf-idf weighting scheme for building features. In addition they built a network composing of bullies and their victims. The network was used to rank the most active predator and its target. In [42] Dinakar et al. stated that detecting profane language cyberbullying is easier than detecting sarcasm and subtle language attacks. Chayan and Shylaja of [43] enhanced the performance of the cyberbullying detection model by 4%, through looking for comments directed towards peers by using Supervised ML and Logical Regression models. However they didn't detect sarcasm comments. Dadvar et al. [44] state that incorporating user context such as the user's history as a feature for training the cyberbullying detection model increases accuracy of classification; however, they didn't include sarcasm detection in their system.

SVM was also used by Yin et al. [45] for classifying posts as containing harassment or not. They used documents from CAW 2 dataset, which included posts from Kongregate, Slashdot and Myspace. For feature selection they incorporated several features;

- *Local features,* they used tf-idf.
- S*entiment features,* such as 1- grams, 2- grams and 3- grams, and they also captured second pronouns.
- *Contextual features* in which they studied both the similarity of a post to other neighboring posts and the cluster of posts neighboring around a certain harassment post.

Capturing sentiment features only didn't perform well so they compared the performance of their system by mixing features. Tf-idf performed better than n-grams and foul words, however, combining tf-idf with contextual and sentiment features achieved an additional enhancement in results in Precision, Recall and F1-measure. A similar work was done by Dadvar et al. [44], who built their feature space from Content-based, User based and Context based features. They also proved that incorporating contextual features such as gender information from the user's profile enhances cyberbullying detection.

Other research efforts were focused around social network profiles, such as [15]. They presented a methodology to detect and associate fake profiles on Twitter social networks to real users. This system had been capable of linking the owners of a fake account on Twitter to a real account for one or more students in a school class; this was a case of a real cyberbullying incident. The system was devised by collecting features from tweets then analyzing the features using various supervised ML techniques included in WEKA. Afterwards the performance among these techniques was compared on True Positive Ratio (TPR), False Positive Ratio (FPR) and Area Under ROC Curve (AUC). Bayzick, Kontostathis and Edwards [46] proposed the BULLYTRACER software which detects cyberbullying in chat rooms 58.63% of the time.

Chen et al. [47] proposed a new model for detection which they named as the Lexical Syntactic Feature-based (LSF) model; it achieved a precision of 98.24% and recall of 94.34%. Their model calculated both a post and a user's offensiveness depending on the ratio of offense appearing in a user's posts. This model detects "strong profanity" in online posts by using lexical analysis methods such as Bag of Words; and subtle language harassment to which the authors referred as "weak profanity". Then the model uses semantic analysis and NLP techniques to analyze the context of sentences by studying the grammatical relations among words. This research was an extension to the work presented in [48] for cyberbullying posts filtration.

Most of the research in cyberbullying did not give importance for the distinction between cyberbullying and cyberagression, but for Hosseinmardi et al. [49]. They proposed a definition for cyberbullying which is the ***repetition*** of harmful actions using electronic devices over a certain period of time. They stated that most of the work previously done in detecting cyberbullying was actually focusing on detecting cyberagression. They define cyberagression as a single instance of harmful action that if repeated over time would be considered as cyberbullying. They also demonstrated that a Linear SVM classifier can significantly improve the accuracy of identifying cyberbullying to 87%. In addition they incorporated using features other than text such as images for better detection of cyberbullying. Potha and Maragoudakis [50] stressed on a window of time in order to study the textual patterns of previous conversations in order of predicting the upcoming actions of a predator. They incorporated time series modelling in their research in addition to SVM for features selection. SVD (Singular Value Decomposition) [51] was used for feature reduction and DTW (Dynamic Time Warping) [52] for matching time series collections.

Fuzzy Logic and Genetic algorithms were also used in cyberbullying detection [53], where a new system was proposed using those two methods. This system's performance was compared across precision, recall and F1 measures. The system achieved better in Accuracy, F1-measure and Recall than previous fuzzy classification methods with 0.87, 0.91 and 0.98 respectively.

B. Arabic Language

Work related to Arabic language is scarce due to the complex morphological nature of Arabic. Arabic language is used by around 300 million Arabs around the world, mainly Muslims. It is a script language which is read and written from right to left and it constitutes of an alphabet of 28 letters. Vowels in Arabic are represented by special punctuation marks called Diacritics [54]. There are three variations for Arabic languages going on together. The Classical Arabic which is the language of the Islamic manuscripts -such as the Quran and prayers- and Arab people until Mid-20th century. The Modern Standard Arabic (MSA) which is the formal language used nowadays in schools and news, and it is known by all Arabs. Finally comes the dialects, which are accents for the Arabic language, usually used informally between people. There are around 10 dialects, one for every country - or group of countries [55], [56]. Arabic dialects imply a difference in meanings of words between different countries. We might find some words that are considered profane in one country, while good or ordinary in others, for example the word "Yetqalash" in Yemen is a compliment while in Morocco it is an offensive word [57].

Arabic language is a challenging and complex language due to its nature, where Arabic words do not include capitalization [56]. The morphologic nature of Arabic inflicts a lot of ambiguity, and the Arabic corpus is very scarce. Arabic language is ranked the 7th around the world, and its use over internet is growing vastly [55], thus arose the research interest in Arabic language fields.

An extensive search was performed on available articles and publications and no previous work for cyberbullying detection in Arabic language texts and comments was found. But some papers in the fields of ML and NLP were found. The previous work done in Arabic deals with text preprocessing or text classification.

Ghaleb Ali and Omar [58], proposed a key phrase extraction method that combines several key phrase extraction methods with ML algorithms. The output from the key phrase extraction methods is used as features to the ML algorithms. The ML algorithm in turn classifies the feature as either a key phrase or not. They compared their results by using three ML algorithms: Linear Logistic Regression [59],

SVM and Linear Discriminant Analysis [60]. They have proved that SVM gives the best results in key phrases extraction among the three algorithms.

Some work had been done in Arabic named entity extraction, such as Named Entity Recognition for Arabic (NERA) [61] to identify proper names in Arabic documents. NERA used a whitelist of named entities and corpora compiled from various sources; its performance was measured across recall, precision and F1. The results were satisfactory; 86.3% 89.2% 87.7% respectively for person named entities.

Filtering for spam emails written in Arabic and English was done by El-Halees [62] on pure English, pure Arabic and mixed collections of emails. Several ML techniques were used, including SVM, NB, k-Nearest Neighbor (k-NN) [63] and Neural Networks. The performance of the system was measured across all three variations and SVM was proved to be best in pure English environment. The system performed less on pure Arabic emails, due to the inflictive nature of Arabic Language. The authors also proved that stemming Arabic words enhances the performance of the classifiers, where NB performed best with 96.78% Recall and 92.42% F1-measure.

Other than emails there are also attempts for detecting spam in social networks, such as Twitter. Such work was done by El-Mawass and Alaboodi in [64]. They elaborated a system to detect spam in Arabic tweets. Their system achieved significant accuracy, precision and recall measures.

Sentiment analysis is one of the text classification categories. Sentiment analysis classifies a certain text as positive, negative or neutral [65]. Sentiment analysis was done by Hamouda [64] on Facebook comments written in Arabic. They built a corpora from 6000 comments sampled from Facebook, preprocessed this corpora, and then applied classifications to determine the sentiment behind the comment. Three classifiers were used: SVM, NB and Decision trees. The best performance was achieved by SVM with 73.4% accuracy. Another attempt for sentiment analysis was done in [65] for Arabic Tweets, their special contribution was in handling Arabizi and dialects. They incorporated NB, SVM and k-NN for classification and the best accuracy was approached by NB.

Sentiment analysis was applied on Arabizi also by Duwairi et al [65]. In their system they first converted Arabizi text into Arabic by using their own rule based method. They labeled their data using their crowdsourcing tool [67] and then applied SVM and NB for classification. A comparison between SVM and NB showed that SVM outperformed NB. However better results were achieved when they first eliminated neutral entries from the dataset.

A significant research effort was done on stemming for Arabic language. Stemming is a text preprocessing technique. In stemming words are truncated to obtain their roots [66]. Several stemmers for Arabic are available, including rule based stemmers such as Khoja's [67] and light stemmers. Light stemmers blindly remove letters from words – affixes and suffices – without prior knowledge of roots [68]. Stemmers are either monolingual or multilingual. Gadri and Moussaoui [69] elaborated a multilingual stemmer. Their stemmer is Language independent and it used the n-gram technique. This stemmer segments words into bigrams, then statistical measures are used to reach the best root. This stemmer was tested against English, French and Arabic. The best success rate (94%) was achieved in small Arabic Datasets. In large datasets, the best results were for English (86, 50%) and the worst for Arabic (67, 66%).

IV. Future Work

Up to the time of writing this paper, we couldn't find any research dealing with Arabic cyberbullying detection; thus we propose a multilingual cyberbullying detection system. This proposed system will benefit from the available ML and NLP techniques. The proposed system will efficiently detect cyberbullying incidents happening on the Online Social Networking (OSN) platforms such as Facebook and Twitter. Our system will tackle cyberbullying content in pure Arabic, pure English and mixed environments which include Arabish or Arabizi texts. We are planning to build our datasets by scrapping comments from the OSN platforms mentioned above, by using some available data management platforms, such as HP Autonomy products - Intelligent Data Operating Layer (IDOL) [70]. We plan to classify our data using several ML techniques from the WEKA toolkit, and we will measure our results for performance across the levels of Recall, precision and F- measure in order to reach a system with optimum performance.

References

[1] K. Poels, A. DeSmet, K. Van Cleemput, S. Bastiaensens, H. Vandebosch and I. De Bourdeaudhuij, "Cyberbullying on social network sites. An experimental study into bystanders," Cyberbullying on social network sites, vol. 31, p. 259–271, 2014.

[2] S. S. Kazarian and J. Ammar, "School Bullying in the Arab World: A Review," The Arab Journal of Psychiatry , vol. 24, no. 1, pp. 37 - 45, 2013.

[3] ICDL, "Cyber Safety Report: Research into the online behaviour of Arab youth and the risks they face," ICDL Arabia, 2015.

[4] K. DINAKAR, B. JONES, C. HAVASI, H. LIEBERMAN and R. PICARD, "Common Sense Reasoning for Detection, Prevention,and Mitigation of Cyberbullying," in ACM Transactions on Interactive Intelligent Systems, NY, September 2012.

[5] O. f. V. o. C. National Crime Prevention Council, "Cyberbullying Tip Sheets," National Crime Prevention Council, 2016. [Online]. Available: http://www.ncpc.org/topics/cyberbullying/cyberbullying-tip-sheets/. [Accessed 10 June 2016].

[6] N. Willard, "Educator's Guide to Cyberbullying and Cyberthreats," Center for Safe and Responsible Internet Use, 2007.

[7] N. Samaneh, A. Masrah, M. Azmi, M. S. Nurfadhilna, A. Mustapha and S. Shojaee, "13th International Conrence on Intelligent Systems Design and Applications (ISDA)," in A Review of Cyberbullying Detection . An Overview, 2013.

[8] D. Mann, "Emotional Troubles for 'Cyberbullies' and Victims," WebMD Health News, 6 July 2010. [Online]. Available: http://www.webmd.com/parenting/news/20100706/emotional-troubles-for-cyberbullies-and-victims. [Accessed 24 August 2015].

[9] T. M. Mitchell, "The Discipline of Machine Learning," CMU-ML-06-108, Pittsburgh, July 2006.

[10] P. Kulkarni, Reinforcement And Systemic Machine Learning For Decision Making, New Jersey: IEEE, WILEY, 2012.

[11] P. FLACH, MACHINE LEARNING The Art and Science of Algorithms that Make Sense of Data, Cambridge University Press, 2012.

[12] D. Vilariño, C. Esteban, D. Pinto, I. Olmos and S. León, "Information Retrieval and Classification based Approaches for the Sexual Predator Identification," Faculty of Computer Science, Mexico.

[13] H. José María Gómez and A. A. Caurcel Diaz, "Combining Predation Heuristics and Chat-Like Features in Sexual Predator Identification," 2012.

[14] A. S. a. S. Vishwanathan, Introduction to Machine Learning, Cambridge: Cambridge University Press, 2008.

[15] I. Santos, P. G. Bringas, P. Gal´an-Garc´ıa and J. Gaviria de la Puerta, "Supervised Machine Learning for the Detection of Troll Profiles in Twitter Social Network: Application to a Real Case of Cyberbullying," DeustoTech Computing, University of Deusto, 2013.

[16] I.-S. Kang, . C.-K. Kim, . S.-J. Kang and S.-H. Na, IR-based k-Nearest Neighbor Approach for Identifying Abnormal Chat Users, 2012.

[17] C. M. a. G. Hirst, Identifying Sexual Predators by SVM Classification with Lexical and Behavioral Features, 2012.

[18] D. E. L. a. A. B. Javier Parapar, "A learning-based approach for the identification of sexual predators in chat logs," 2012.

[19] Ron Kohavi and R. Quinlan, "Decision Tree Discovery," 1999.

[20] K. Reynolds, "Using Machine Learning to Detect Cyberbullying," 2012.

[21] S. Ahmad, "Tutorial on Natural Language Processing," Artificial Intelligence (810:161) Fall 2007.

[22] V. Gupta, "A Survey of Natural Language Processing Techniques," vol. 5, 01 Jan 2014.

[23] B. MANARIS, "Natural Language Processing: A Human–Computer Interaction Perspective," vol. 47, no. pp. 1-66, 1998..

[24] E. Cambria and B. White, "Jumping NLP Curves: A Review of Natural Language Processing Research," IEEE ComputatIonal IntEllIgEnCE magazIne, May 2014.

[25] C. Surabhi.M, "Natural Language Processing Future," in International Conference on Optical Imaging Sensor and Security, Coimbatore, Tamil Nadu, India, July 2-3, 2013.

[26] G. G. Chowdhury, "Natural Language Processing," Annual Review of Information Science and Technology, vol. 37, no. 0066-4200, pp. 51-89, 2003.

[27] E. Cambria, Application of Common Sense Computing for the Development of a Novel Knowledge-Based Opinion Mining Engine, University of Stirling, Scotland, UK, 2011.

[28] M. Grassi, E. Cambria, A. Hussain and F. Piazza, "Sentic Web: A New Paradigm for Managing Social Media Affective Information," Cogn Comput (2011) 3:480–489.

[29] W. E. Webber, Measurement in Information Retrieval Evaluation (Doctor of Philosophy), The University of Melbourne, September 2010.

[30] C. J. v. RIJSBERGEN, INFORMATION RETRIEVAL, University of Glasgow.

[31] N. Chinchor, "MUC-4 EVALUATION METRICS," in Fourth Message Understanding Conference, 1992.

[32] Y. Sasaki, "The truth of the F-measure," University of Manchester, 26th October, 2007.

[33] "Arabic chat alphabet," 23 May 2016. [Online]. Available: https://en.wikipedia.org/wiki/Arabic_chat_alphabet. [Accessed 2 June 2016].

[34] WatchGuard, "Stop Cyber-Bullying in its Tracks - Protect Schools and the Workplace," WatchGuard Technologies, 2011.

[35] "https://blog.barracuda.com/2015/02/16/3-ways-the-barracuda-web-filter-can-protect-your-classroom-from-cyberbullying/".

[36] "Internet Monitoring and Web Filtering Solutions," PEARL SOFTWARE, 2015. [Online]. Available: http://www.pearlsoftware.com/solutions/cyberbullying-in-schools.html. [Accessed 2 June 2016].

[37] V. Nahar, X. Li and C. Pang, "An Effective Approach for Cyberbullying Detection," in Communications in Information Science and Management Engineering, May 2013.

[38] "Perverted Justice," Perverted Justice Foundation, [Online]. Available: http://www.perverted-justice.com/.

[39] "Amazon Mechanical Turk," 15 August 2014. [Online]. Available: http://docs.aws.amazon.com/AWSMechTurk/latest/AWSMechanicalTurkGettingStartedGuide/ SvcIntro.html. [Accessed 2 June 2016].

[40] S. Garner, "Weka: The waikato environment for knowledge analysis," New Zealand, 1995.

[41] "tf-idf: A single Page Tutorial," [Online]. Available: http://www.tfidf.com. [Accessed 13 May 2016].

[42] K. Dinakar , R. Reichart and H. Lieberman, "Modeling the Detection of Textual Cyberbullying," Cambridge, 2011.

[43] V. S. Chavan and Shylaja S S , "Machine Learning Approach for Detection of Cyber-Aggressive Comments by Peers on Social Media Network," in International Conference on Advances in Computing, Communications and Informatics (ICACCI), 2015.

[44] M. Dadvar, D. Trieschnigg, R. Ordelman and F. De Jong, "Improving cyberbullying detection with user context," 2013.

[45] D. Yin, Z. Xue, L. Hong, B. D. Davidson, A. Kontostathis and L. Edwards, "Detection of Harassment on Web 2.0," Madrid, Spain, April 21, 2009.

[46] J. Bayzick, A. Kontostathis and L. Edwards, "Detecting the Presence of Cyberbullying Using Computer Software," Koblenz, Germany, June 14-17, 2011.

[47] Y. Chen, S. Zhu, Y. Zhou and H. Xu, "Detecting Offensive Language in Social Media to Protect Adolescent Online Safety," 2012.

[48] Z. Xu and S. Zhu, "Filtering Offensive Language in Online Communities using Grammatical Relations," Redmond, Washington, US, July 13-14, 2010.

[49] H. Hosseinmardi, S. Arredondo Mattson, R. IbnRafiq, R. Han, Q. Lv and S. Mishra, "Detection of Cyberbullying Incidents on the Instagram Social Network," 2015.

[50] N. Potha and M. Maragoudakis, "Cyberbullying Detection using Time Series Modeling," 2014.

[51] K. Baker, "Singular Value Decomposition Tutorial," 2013.

[52] M. Muller, "Dynamic Time Warping," in Information Retrieval for Music and Motion, Springer, 2007, pp. 69 - 84.

[53] B. Nandhinia and J. Sheebab , "Online Social Network Bullying Detection Using Intelligence Techniques," 2015.

[54] M. A. Attia, Handling Arabic Morphological and Syntactic Ambiguity within the LFG Framework with a View to Machine Translation, Doctor of Philosophy in the Faculty of Humanities, 2008.

[55] K. Darwish and W. Magdy, "Arabic Information Retrieval," vol. 7, no. 4, 2013.

[56] A. FARGHALY and K. Shaalan, "Arabic Natural Language Processing:Challenges and Solutions," vol. 8, December 2009.

[57] "12 Arabic Swear Words and Their Meanings You Didn't Know," [Online]. Available: http://scoopempire.com/swear-words-meanings-around-middle-east/#.V0fdjPl96M9. [Accessed 2 June 2016].

[58] N. Ghaleb Ali and N. Omar, "Arabic Keyphrases Extraction Using a Hybrid of Statistical and Machine Learning," in International Conference on Information Technology and Multimedia (ICIMU), Putrajaya, Malaysia, 2014.

[59] T. Haifley, "Linear Logistic Regression: An Introduction," IEEE, 2002.

[60] G. J. McLACHLAN, "Discriminant Analysis and Statistical Pattern Recognition," Wiley InterScience, New Jersey, 2004.

[61] K. Shaalan and H. Raza, "Arabic Named Entity Recognition from Diverse Text Types," Berlin Heidelberg, GoTAL 2008.

[62] A. El-Halees, "Filtering Spam E-Mail from Mixed Arabic and English Messages: A Comparison of Machine Learning Techniques," The International Arab Journal of Information Technology, vol. 6, no. 1, 2009.

[63] T. M. COVER and P. E. HART, "Nearest Neighbor Pattern Classification," IEEE TRANSACTIONS ON INFORMATION THEORY, vol. 13, no. 1, 1967.

[64] A. E.-D. A. Hamouda and F. E.-z. El-taher, "Sentiment Analyzer for Arabic Comments System," (IJACSA) International Journal of Advanced Computer Science and Applications, vol. 4, no. 3, 2013.

[65] R. M. Duwairi, R. Marji, N. Sha'ban and S. Rushaidat, "Sentiment Analysis in Arabic Tweets," in 5th International Conference on Information and Communication Systems (ICICS), 2014.

[66] A. Al-Zyoud and W. A. Al-Rabayah, "Arabic Stemming Techniques: Comparisons and New Vision," in Proceedings of the 8th IEEE GCC Conference and Exhibition, Muscat, Oman, 2015.

[67] S. Khoja and R. Garside, "Stemming arabic text," Computing Department, Lancaster University, Lancaster, UK, 1999.

[68] L. S. Larkey, L. Ballesteros and M. E. Connell, "Light Stemming for Arabic Information Retrieval," in Arabic Computational Morphology, book chapter, , , Springer, 2007.

[69] S. Gadri and A. Moussaoui, "Information Retrieval: A New Multilingual Stemmer Based on a Statistical Approach," in 3rd International Conference on Control, Engineering & Information Technology (CEIT), 2015.

[70] Hewlett-Packard Development Company. L.P., 2013. [Online]. Available: http://www.autonomy.com/html/power/idol-10.5/index.html. [Accessed 2 June 2016].