A review of cyberbullying detection: An overview

Conference Paper · December 2013

DOI: 10.1109/ISDA.2013.6920758

Samaneh Nadali
Universiti Putra Malaysia
5 PUBLICATIONS **59** CITATIONS
SEE PROFILE

Masrah Azrifah Azmi Murad
Universiti Putra Malaysia
152 PUBLICATIONS **660** CITATIONS
SEE PROFILE

Nurfadhlina Bt Mohd Sharef
Universiti Putra Malaysia
78 PUBLICATIONS **193** CITATIONS
SEE PROFILE

Aida Mustapha
Universiti Tun Hussein Onn Malaysia
178 PUBLICATIONS **545** CITATIONS
SEE PROFILE

A Review of Cyberbullying Detection : An Overview

Samaneh Nadali*, Masrah Azrifah Azmi Murad, Nurfadhlina Mohamad Sharef, Aida Mustapha,Somayeh Shojaee
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
Serdang, Selangor, Malaysia
sm.nadeali@gmail.com*,{masrah, nurfadhlina, aida}@upm.edu.my, somayeh.shojaee@gmail.com

***Abstract*—With the growth of Web 2.0, online communication and social networks are emerging. This alternation helps users to share their information and collaborate with each other easily. In addition, these internet services help establish new connections between persons or reinforce existing ones. However, they can also lead to misbehaviors or cyber criminal acts for example, cyberbullying. At the same time, it can make children and adolescents to use the technologies for the intention of harming another person. Due to the negative effect of cyberbullying, some techniques and methods are proposed to overcome this problem. This paper illustrates a survey covering some methods and challenges in cyberbullying. Next, we offer suggestions for continued research in this area.**

Keywords- cyberbullying; cybercrime; cyber predator; text mining.

I. INTRODUCTION

Nowadays, people around the world use different online forums, blogs, social networking sites, forums as a foundation of their networking, sharing and transfer of knowledge. Although, online communities and social networks have become more common, some users use these communities in illegal and unethical ways, which lead to teens and youth people bullied over the internet.

National Crime Prevention Council delimit cyberbullying as the following: 'when the Internet, cell phones or other devices are used to send or post text or images intended to hurt or embarrass another person' [22]. Dana Boyd as a social scientist has expressed four phrases on the web. These phrases change the very dynamics of bullying. They also magnificent bullying to new levels : search ability, replicability, persistence and invisible audiences [23].

Recent studies reported almost 43% of adolescents in the United States alone been bullied at some point in time [20]. Like traditional forms of bullying, cyberbullying has an intensely negative effect on children and young adults. According to the American Academy of Child and Adolescent Psychiatry, victims of cyberbullying usually suffer from emotional and psychological experience [21].

Social science has been studied extensively for understanding various attributes and the prevalence of cyberbullying problem. Prevention measures include human interference, delete offensive terms, blacklist or scoring of the authors' cyber performance, and educational awareness. However, due to the lack of existing datasets, few studies have been concentrated on online cyberbullying detection. The main course of action in fighting cyberbullying is detection and provision of subsequent preventive measures.

The issues in preventing cyberbullying are finding cyberbullying when it happens; revealing it to police agencies, internet service suppliers and others (for the object of avoidance, education and awareness); and distinguish between predator and victims [24].

In this paper we are going to review some literatures in cyberbullying to know when it occurs and to identify between predators and victim.

This paper is structured into the followings: section 2 describes cybercrime and illustrate two aspects of it i.e. cyberbullying and cyber predator detection. Section 3 presents the data source used for cyberbullying and cyber predator. Section 4 describes some applications and tools that are used in this area. Last section concludes our review and discusses some future direction for research.

II. CYBERCRIME ANALYSIS

Cybercrime is referred to as any illegal activity which make using a computer as the primary means of commission. This definition was expended by the U.S. Department of Justice, for any illegal activity to use a computer as a storage of evidence.

Based on [25], cyber crime could be categorized in two ways: content based and technology based crimes. The former is managed by any specific terrorist organization related to the article of threating , national security, child pornography, sexual harassment, etc. and the latter involves hacking, injecting malicious code, incidents of espionage, etc. The people who are involved in both types should have some technology knowledge. The cyber criminals tend to be residing in various types of world and enjoy getting the privilege of various citizens.

In this paper we consider the content based crime which includes cyber predator and cyberbullying detection.

A. Cyberbullying Detection

Cyberbullying was defined by Patchin and Hinduja as "willful and repeated harm inflicted through the medium of electronic text [3]." From the overview of the adolescent psychology literature shows nine various kinds of cyberbullying which can be recognized in [2], [3], [4]. These categories are: flooding, masquerade, flaming, trolling, harassment, cyberstalking, denigration, outing, and exclusion. The types of bullying are defined as follows :

978-1-4799-3516-1/13/$31.00 ©2013 IEEE

Flooding consists of the bully frequently sending the same comment, nonsense comments, or press the enter key in order to not allow the victim to contribute to the conversation [2].

Masquerade involves the bully pretends to be someone who they are not. This would make it appear with the purpose of bully a victim directly [4].

Flaming, or bashing is a kind of online fight. The bully sending or posting electronic message which are enticingly insulting , vulgar to one or several persons either privately or publicly to an online group [4].

Trolling, also called baiting, includes purposely publishing comments which disagree with other comments. The poster intended to incite an argument or arouse emotions, however the comments are not necessarily personal, vulgar or emotional [1].

Harassment is the kind of conversation that the bully frequently sends insulting and rude messages to the victim [4].

Cyberstalking and cyberthreats occurs when the poster sends intimidating or offensive messages [4].

Denigration also called “dissing” happens when an electronic bully sends or publishes gossip or untrue statement about a victim in order to damage the victim’s friendship or reputation [4].

Outing occurs when a person sends or publishes private or embarrassing information in a public chat room or forum. This type of cyberbullying is similar to the denigration. However in outing the relationship between bully and victim are close [4].

Exclusion involves intentionally excluding someone from an online group. This type of cyberbullying happened among youth and teenage generally [3].

For a number of issues related to cyberbullying recognition, research has been done based on the text mining paradigm such as online sexual predator recognition [7] and spam detection [8]. Nevertheless, very little study has been done on technical solutions, for which is why there is insufficient proper training datasets. Moreover, privacy issues and ambiguities can be the reasons in describing cyberbullying.

References [5] proposed a supervised learning approach for determining harassing posts in chat rooms and discussion forums. Three types of features, namely content, sentiment and contextual were used for training a support vector machine classifier. They also used N-grams, TFIDF weighting and foul word frequency as the base-line. Although their results indicate enhancements over the baselines, the temporal or user information have not been utilized. Moreover they employed only supervised methods. Nevertheless, unsupervised methods may also prove to be valuable. In another study with the same dataset the authors tried to identify clusters containing cyberbullying using a rule-based algorithm [9].

For detecting cyberbullying among YouTube comments, researchers [6] described a method. In their method, they used a variety of binary and multiclass classifier on a manually labelled dataset. Also they applied common sense knowledge for detecting cyberbullying. Using common sense can help provide information about people’s goals and emotions and object’s properties and relations that can help disambiguate and contextualize language. They also used two types of features: 1) general features that contain a term frequency-inverse document frequency (TF-IDF) weighted uni-grams, the Ortony lexicon of words denoting negative connotation, a list of profane words and frequently occurring part-of-speech (POS) bigram tags and 2) label specific features. Their study indicated that binary classifier can outperform the recognition textual cyberbullying in comparison to multiclass classifiers. Their results illustrate using such features into account will be more useful and can lead to better modelling of the problem. The limitations of their study are that they did not consider the pragmatics of dialogue and conversation and the social networking graph.

Reference [14] improved the work that was done by Massachusetts Institute of Technology’s approach [6]. They proposed a machine learning approach for detecting cyberbullying from Formspring.me. They applied the number of "bad" words (NUM) and the density of "bad" words (NORM) features that were devised by assigning a severity level to the badwords list (nosewaring.com). They employed replication of positive examples up to ten times and accuracy on the range of classifiers was reported. Their results illustrated that the C4.5 decision tree and an instance based learner could recognize the true positives by 78.5% accuracy.

Recently some works have been done on recognizing users by interaction on the web. While providing profile information for social networks, browsing the internet, users leave an amount of traces. This distributed user data can be utilized as a way to obtain information for systems that provide personalized services for their users or need to find more information about their users [12]. Connecting data from different sources has been used for different purposes, such as standardization of APIs (e.g. OpenSocial1) and personalization [10].

Previous works in cyberbullying detection have mostly concentrated on the conversations' content though they did not attend to the characteristics of the actors involved in cyberbullying. Social studies demonstrate that male and female bully each other in different way. For example, while women with aggressive communication styles, such as excluding someone from a group of conspiracy against them, men tend to use more words and phrases threatened outrage [11]. Reference [26] illustrates that more pronouns like “I”, “you”, “she”, etc. are used by females and more noun specifiers such as ,“a”, “the”, “that” are used by males. These findings have motivated a recent study [13] of the effect of linguistic features based on gender in the diagnosis of cyber bullying on social networks.

Reference [24] in 2013, proposed an effective approach to detect cyberbullying from social media. And they also presented a graph model to extract cyberbullying network. This has led to identifying the most active predators and victims through a ranking algorithm. Their proposed graph model could be used to recognize the level of cyberbullying victimization for decision making in further studies. They could improve the classification performance by applying a weighted TF-IDF function, in which bullying-like features are scaled by a factor of two.

As we have mentioned before, there are four major tasks in cyberbullying detection: detecting online bullying; reporting it to law enforcement agencies, Internet service providers and others (for the purpose of prevention, education and awareness); and identifying predators and their victims. In part A, we review some techniques in cyberbullying detection. Next session we describe some papers in detection cyber predators and victim detection.

B. *Cyber predator*

> "A cyber predator is a person who uses the Internet to hunt for victims to make the most of in any way, including sexually, emotionally, psychologically or financially. Cyber predators know how to manipulate kids, creating trust and friendship where none should exist" [27].

National Center for Missing and Exploited Children (NCMEC), explained about 1 in 7 youth (ages10- to 17-years-old) experience a sexual approach or appeal through the Internet [18].

Online sexual predator studies [16], [17] relate to the theory of communication and text-mining techniques to distinguish between predator and victim conversations, as applied to one-to-one communication.

Recognizing the predator problem is divided into two sub problems, namely identifying predators and recognizing predators' lines for identifying predators. Based on [39] we can divide the summary of existing approaches in identifying predators into three steps namely: pre-filtering approach, feature extraction approach, and classification approach. One of the most effective methods for pre-filtering of all conversations is done by [31]. They displayed some specific patterns, for example, the existence of one participant only, those with less than 6 interventions per user or those which included 3 long sequences of unrecognized characters. Although other researchers [38] proposed similar task, they applied a rule based approach on different features for different methods.

For the second task i.e. feature extraction, the features are categorized into two principal groups: "lexical" features and "behavioral" features. The former are those which can be derived from the raw text of the conversation, for example unigram or bigram [31,37,35,38] features, emoticons counting and the weighting applied TF-IDF or the cosine similarity. Recognizing the name of the participants in the conversation (self, other, group) is an another example [35].

The latter features are within a conversation [43,36] : the number of questions asked , intention (grooming, hooking, ...), its capture the "action" of the users. The creation of single set features for each author is one of the important approaches. This approach can describe and develop his predator potential. Some researchers used the Language Model (LM) for two participants in the chat [35]. There are some approaches which used LM at line level or at conversation level. They applied this strategy to sum up the score of all the lines or conversation to find a unique set of features of each author [41,42,43,40,45].

For classifying predator and non-predator many different approaches were proposed, for example, decision trees [42], random forest [46] as well as Naïve Bayes [43,41] and Maximum-Entropy [35,45]. In comparison among existing classifier, Support Vector Machines (SVM) were used a lot [37,38,40,31]. Some researchers have shown the other approach performed better than SVM, for instance, when they applied a Neural Network classifier [31].

For the second issue which is identifying predator's lines, the proposed solution was related to all the relevant conversation lines of all predators. These predators are obtained from the first problem [46]. One of the most used technique was a filtering of all the conversations of predator via a thesaurus of "perverted" phrases or with a specific score (e.g. TF-IDF weighting) [40,37,38,35].

The final approach was simply to return those lines previously labelled as predatory in the proposed algorithm by the default method for working at line level [42,43,45].

III. Data source

In this section we present an overview of the collection and labelling the data which are used in previous approaches.

A) *Dataset Origin*

In predator communications there is very little reliable labelled data; where a lot of the work in both communication studies and computer science is dedicated to anecdotal evidence and chat logs transcripts from Perverted Justice (PJ)[47].

Using the PJ transcript for cyber predator detection is contentious. The PJ contain transcripts of conversation. These conversations are involved between a predator and pseudo-victim, an adult posing as a youth.

Dr. Susan Gauch, University of Arkansas made the second data set to recognize a predator [48]. She developed a new software that called ChatTrack for crawling and downloading chat logs. Although, ChatTrack is not accessible now, the chart data are still used in some of the primary research. The researchers have included analyses of predator communication [15].

In the Content Analysis for Web 2.0 workshop (CAW 2.0-2009) a shared task for misbehavior detection was proposed. Misbehavior detection is to recognize improper activity in virtual community when some users harass or offend other members. CAW 2.0 provides datasets for online harassment. The dataset contains five various public sites; Ciao, MySpace, Twitter and Kongregate [19]. These data can be categorized into two kinds of communities; chat-style and discussion-style communities. Among the mentioned datasets, Kongregate is one of the samples of chat-style communities. In this type of dataset, posts are contained of short messages. The main characteristics of these messages are having a few words with many misspellings. In comparison with chat-style communities, discussion-style communities (MySpace and Slashdot) have rather longer posts. These posts are still shorter than full web pages. The terms in these posts are also more formal.

In chat-style communities like Kongregate, posts are usually short online messages, which contain only a few words with many misspellings. In discussion-style communities, like Slashdot and MySpace, posts are relatively longer (but still shorter than full web pages) and the usage of the terms in these posts is more formal, as compared with other chat-style communities.

B) Labeling the Data

For generating a coded datset, few numbers of devoted content coders were employed by previous researchers. Amazon Mechanical Turk (MTurk) has been used by recent researchers as the crowdsourcing services. MTurk is an online labor market which helps researchers to decompose a request post jobs into a large amount of small tasks. Tukers (MTurk workers) have offered a brief description of accessible tasks and determine which task to accomplish. Usually used time for accomplishing each task is between 5 and 20 seconds. Also workers paid about 5 cents for every single task. Besides its function as a source of labor, MTurk is a good place to conduct user studies in human-computer interaction (Kittur, Chi, & Suh, 2008) and large-scale economic experiment (Mason & Watts, 2009).

In comparison between traditional model , MTurk has several significant advantages. Firstly, Turkers are an on-demand labor force. Using the MTurk system we could also rapidly and engaged a large number of programmers, without the significant overhead associated with hiring dedicated employees . Secondly, MTurk workers are derived from a various perspective and provide several different experience with online interaction and comments.

An important factor of MTurk is the quality. Some studies [49] illustrated utilizing MTurk for analyzing content was faster and cheaper than utilizing devoted raters. Other researchers [50] also demonstrated that using various non expert workers could make high quality results.

Lastly, [51] proposed which it is possible to obtain high quality, efficient coding through the use of several non-expert coders even though individual coders do not always agree (i.e. the data are “noisy”).

IV. Application and tools

With the growth of cyberbullying among children and teenagers, the most important question that can be expected from a teenager is to discern the degree between right and wrong? So responsible parents need to protect their children from internet predators. In this regards some available commercial and networks are eBlaster TM, Net Nanny TM, and , IamBigBrother [32,33,34] .

Packet sniffing is the most prevalent alternative to Safe Chat. Packet sniffers scan all the outgoing and ingoing traffic in a network and then apply a filter to only see the useful blocks of data. Although many packet sniffers and tools are easy to use , parents may have problem with these tools. The most important reason is that tools which are now available to detect predation are based on a simple keyword matching and not communication theory. This brings the accuracy of these tools into question [15].

In order to overcome the limits of other tools , SafeChat was proposed. This software is also better than currently available tools. The first version of SafeChat was stand-alone software. SafeChat 1.0 used the WinpCap library. WinpCap is a library that helps programmers to have high level control over the retrieval and transmission of packets in the Windows environment. This library is used by many widely used commercial products such as Wireshark [29].

SafeChat 1.0 was designed to work with AIM Instant Messaging because AIM has the largest market share among IM tools. AIM uses a protocol called Open System for Communication in Realtime (OSCAR). Despite the name, OSCAR is not an Open Source System. In 2008, documentation on the OSCAR protocol was released [28]. Like AIM, many other chat clients did not have proper documentation. SafeChat, to be a successful, has to be compatible with many protocols.

SafeChat 2.0 is the new version of SafeChat. It is a third party plugin for the Pidgin, an open source instant messaging system. It uses detection algorithms to classify chat participants as potential predators.

Pidgin is one of the most popular open source instant messaging systems. It works on any Windows or Unix-based environment and supports multiple protocols including AIM, MSN, ICQ, IRC, and Yahoo. Unsupported protocols, like Facebook Chat, can be used in Pidgin with the use of third party plugins [30]. There are multiple reasons for choosing the Pidgin platform. The primary reason is that we want SafeChat to be available to assist as many families as possible. Therefore, SafeChat needs to support as many IM protocols as possible. Second, SafeChat can take advantage of the development efforts of the Pidgin community. When new protocols are made or existing protocols are changed, the Pidgin community will update Pidgin. This allowed us to focus on the predation algorithms for SafeChat instead of on infrastructure issues.

V. Conclusion

With the rapid growth of the internet, more and more people interact with other people in the same town or the other side of the world. However, the chance for misuse comes with any new technology. Unfortunately, these techniques lead to misbehavior or cyber criminal act like cyberbullying. Our literature review illustrates that there are few research on cyber predator and cyberbullying detection. In the future, addressing the role of newer technologies especially peer-to-peer device and cell phones should be considered for further researches.

Also, collaborations with text mining and information retrieval research groups help us to find a good solution to detect this annoying problem. As we mentioned before, there is no labelled dataset, so in future researcher can work on collecting new labelled dataset for the future study. Working with psychologists, sociologists, communications and law enforcement experts can improve awareness of understanding, recognizing and preventing cyber crime. And developing a good classifier to recognize predator’s behavior is also needed.

REFERENCES

[1] Glossary of cyberbullying terms. (2008, January). Retrieved from http://www.adl.org/education/curriculum_connections/cyberbullying/glossary.pdf

[2] D. Maher, "Cyberbullying: an ethnographic case study of one australian upper primary school class". Youth Studies Australia, 27(4), 5057,2008.

[3] J. Patchin , & S. Hinduja, "Bullies move beyond the schoolyard; a preliminary look at cyberbullying." Youth violence and juvenile justice. 4:2 (2006). 148-169.

[4] N.E. Willard, "Cyberbullying and Cyberthreats: Responding to the Challenge of Online Social Aggression, Threats, and Distress." Champaign, IL: Research, 2007.

[5] D. Yin, Z. Xue, L. Hong, B.D. Davison, A. Kontostathis, L. Edwards, " Detection of harassment on Web 2.0." In: Proceedings of CAW2.0, 2009.Madrid, April 20-24.

[6] K. Dinakar, R. Reichart, H. Lieberman, "Modelling the Detection of Textual Cyberbullying." In: ICWSM 2011, Barcelona, Spain, July 17-21 2011.

[7] A. Kontostathis, "ChatCoder: Toward the tracking and categorization of internet predators." In: Proceedings of SDM 2009, Sparks, NV, May 2 2009.

[8] P.N. Tan,F. Chen, A. Jain, "Information assurance: Detection of web spam attacks in social media." Proceedings of Army Science Conference, Orland, Florida. 2010.

[9] J.F. Chisholm, "Cyberspace violence against girls and adolescent females." Annals of the New York Academy of Sciences 1087, 2006. pp. 74–89.

[10] F. Nola, F. Cena, "User identification for cross-system personalisation. Information Sciences" 179, 2009. pp. 16–32.

[11] J.F. Chisholm, "Cyberspace violence against girls and adolescent females." Annals of the New York Academy of Sciences 1087,2006. pp. 74-89.

[12] F. Abel, N. Henze, E. Herder, D. Krause, "Linkage, aggregation, alignment and enrichment of public user profiles with Mypes." In: Proceedings of I-SEMANTICS, Graz, Austria. 2010. pp. 1–8

[13] M. Dadvar, F. d. Jong, R. Ordelman, and D. Trieschnigg, "Improved cyberbullying detection using gender information," In Proceedingsof the Twelfth Dutch-Belgian Information Retrieval Workshop (DIR 2012), February 2012. pp. 23-25,.

[14] K. Reynolds, A. Kontostathis, and L. Edwards, "Using Machine Learning to Detect Cyberbullying," In Proceedings of the 2011 10thInternational Conference on Machine Learning and Applications Workshops (ICMLA 2011), vol. 2, December 2011. pp. 241-244,.

[15] A. Kontostathis, L. Edwards, A. Leatherman, "Text mining and cybercrime." *Text Mining: Applications and Theory. John Wiley & Sons, Ltd, Chichester, UK.*2009.

[16] A. Kontostathis, L. Edwards, and A. Leatherman, "ChatCoder: Toward the Tracking and Categorization of Internet Predators," InProceedings of Text Mining Workshop 2009 held in conjunction with the Ninth SIAM International Conference on Data Mining (SDM 2009) 2009.

[17] I. Mcghee, J. Bayzick, A. Kontostathis, L. Edwards, A. Mcbride, and E. Jakubowski, "Learning to Identify Internet Sexual Predation,"International Journal on Electronic Commerce 2011, vol. 15, pp. 103-122, 2011.

[18] "NCMEC. National center for missing and exploited children", [online] October2008, http://www.missingkids.com/en_US/documents/CyberTiplineFactSheet.pdf.

[19] Fundaci´on Barcelona Media (FBM). Caw 2.0 training datasets. [online] 2009, http://caw2.barcelonamedia.org.

[20] M. Ybarra, "Trends in technology-based sexual and non-sexual aggression over time and linkages to non-technology aggression." Presentation at the National Summit on Interpersonal Violence and Abuse Across the Lifespan: Forging a Shared Agenda. Houston, TX.2010.

[21] "Facts for families, the American Academy of Child Adolescent Psychiatry", [online], http://www.aacap.org/galleries/ FactsFor-Families/80_bullying.pdf.

[22] "Cyberbullying, The National Crime Prevention", [online] http://www.ncpc.org/cyberbullying

[23] D. Boyd, "Why youth (heart) social network sites: The role of networked publics in teenage social life." 2009.

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

[25] M. Thangiah, S. Basri, S. Sulaiman, "A framework to detect cybercrime in the virtual environment". Computer & Information Science (ICCIS), 2012 International Conference on, 2012.

[26] S. Argamon, M. Koppel, J. Fine, , A. R. Shimoni , "Gender, genre, and writing style in formal written texts". Text-The Hague Then Amsterdam Then Berlin-, *23*(3), 321-346,2003.

[27] [online], http://cybersafety.wikispaces.com

[28] [online], http:/dev.aol.com/aim

[29][online],http://web.archive.org/web/20080308233204/http://dev.aol.com/aim/oscar

[30] [online], http://pidgin.im

[31] E. Villatoro-Tello, A. Juárez-González, H. J. Escalante, M. Montes-y-Gómez, and L. V. Pineda, "A Two-step Approach for Effective Detection of Misbehaving Users in Chats" . In CLEF (Online Working Notes/Labs/Workshop) 2012.

[32] eBlasterTM2008. [online], http://www.eblaster.com/

[33] Net NannyTM2008. [online], http://www.netnanny.com/

[34] IamBigBrother n.d. [online], http://www.iambigbrother.com/

[35] G. Eriksson, and J. Karlgren, " Features for modelling characteristics of conversations: Notebook for PAN at CLEF 2012." In CLEF 2012 Evaluation Labs and Workshop Online Working Notes. 2012.

[36] L. Gillam, and A. Vartapetiance, "Quite Simple Approaches for Authorship Attribution, Intrinsic Plagiarism Detection and Sexual Predator Identification".notebook for pan at clef 2012.

[37] C. Morris, and G. Hirst, "Identifying sexual predators by svm classification with lexical and behavioral features" - notebook for pan at clef 2012.

[38] J. Parapar, D.E. Losada, A. Barreiro, "A learning-based approach for the identification of sexual predators in chat logs" - notebook for pan at clef 2012.

[39] G. Inches , and F. Crestani, "Overview of the international sexual predator identification competition at PAN-2012." In CLEF 2012 Evaluation Labs and Workshop — Working Notes Papers. Rome, Italy, 2012.

[40] Peersman, C., Vaassen, F., Asch, V.V., Daelemans, W.: Conversation level constraints on pedophile detection in chat rooms - notebook for pan at clef 2012.

[41] D.V. Ayala, E. Castillo, D. Pinto, I. Olmos, and S. León, "Information retrieval and classification based approaches for the sexual predator identification" - notebook for pan at clef 2012

[42] A. Kontostathis, W. West, A. Garron, K. Reynolds, and L. Edwards, "Identify predators using chatcoder 2.0" - notebook for pan at clef 2012.

[43] J.M.G. Hidalgo, and A.A.C. Díaz, "Combining predation heuristics and chat-like features in sexual predator identification" - notebook for pan at clef 2012.

[44] I.S. Kang, C.K. Kim, S.J. Kang, and S.H. Na, "Ir-based k-nearest neighbor approach for identifying abnormal chat users" - notebook for pan at clef 2012.

[45] R. Kern, S. Klampfl, and M. Zechner, "Vote/veto classification, ensemble clustering and sequence classification for author identification" - notebook for pan at clef 2012.

[46] M. Popescu, and C. Grozea, "Kernel methods and string kernels for authorship analysis "-notebook for pan at clef 2012.

[47] [online], http://www. Perverted-Justice.com .2008.

[48] J. Bengel, S. Gauch, E. Mittur, and R. Vijayaraghavan," ChatTrack: Chat room topic detection using classification". In Intelligence and Security Informatics, pp. 266-277. Springer Berlin Heidelberg, 2004.

[49] J. R. Tetreault, E. Filatova, and M. Chodorow, "Rethinking grammatical error annotation and evaluation with the Amazon Mechanical Turk." In Proceedings of the NAACL HLT 2010 Fifth Workshop on Innovative Use of NLP for Building Educational Applications, pp. 45-48. Association for Computational Linguistics, 2010.

[50] C. Callison-Burch, "Fast, cheap, and creative: evaluating translation quality using Amazon's Mechanical Turk". In Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 1-Volume 1.pp. 286-295. Association for Computational Linguistics ,2009.

[51] V.S. Sheng, F. Provost, and P.G. Ipeirotis, "Get another label? improving data quality and data mining using multiple, noisy labelers." In Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 614-622. ACM, 2008.