Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 72 (2015) 201 – 208

Procedia Computer Science

The Third Information Systems International Conference

Sentiment Valences for Automatic Personality Detection of Online Social Networks Users using Three Factor Model

Saravanan Sagadevan[a], Nurul Hashimah Ahamed Hassain Malim[a*] and Mohd Heikal Husin[a,]
[a]School of Computer Sciences, Universiti Sains Malaysia, Penang, Malaysia.
[a*]Corresponding Author : PH : +604 6534645; Fax : +604 6533335
nurulhashimah@usm.my

Abstract

The interactions using text or writing has become an important medium to communicate among human. The inventions of online networks technologies and applications such as social networks rapidly growth the size of digital textual data and indirectly raises the curiosities to mine the juicy information encapsulated in the text data. Researchers have stated that linguistics of the people either in the online or offline having strong correlations with their self-disclosure. The mass creations of digital textual data raised the intentions of scholars to study the hidden abstractions of personality by applying automatic personality detection approaches. Most of the current automatic personality detection studies focused on Big 5 personality model as a framework to study the underlying characteristics of human. As such, this study incorporates the Three Factor Personality (PEN) Model as a personality framework to guide our understanding and revealing the role of words in depicting the characteristics of a user. This preliminary study revealed how the sentiment perceptions of public towards specific words could assist us in detecting the personality of Facebook users by exploiting their status messages. As the first phase of our study, this experiment focuses on gathering the general perceptions of Malaysians towards 52 English words and 17 interjections that was retrieve from the domain stated above. A Likert scale questionnaire executed to find the sentiment valences of 67 words through analyzing responses from participants and statistical significance will assist the categorization of words under PEN traits. The evaluation provides the necessary analysis that could assist our main research that focuses on automatic personality detections specifically on the psychoticism trait. Our initial findings has highlighted that the five words categorized under psychoticism has strong Cronbach's Alpha coefficients and significant effect from multivariate analysis that indirectly affirmed the reliability of the categorization of the words.

Peer-review under responsibility of organizing committee of Information Systems International Conference (ISICO2015)

Keywords: Sentiment valence; Three Factor Model; Automatic Personality Perception.

1. Introduction

Since the times of Great Greek Philosopher Aristotle, the personality aspect had been studies to understand the thoughts, behaviours and acts of human being. One of the notable and pioneer study that

1877-0509
Peer-review under responsibility of organizing committee of Information Systems International Conference (ISICO2015)
doi:10.1016/j.procs.2015.12.122

described the personality types was found in the book written by one of Aristotle's student, Theophrastus (371 - 287 BC) [3]. Since then, philosophers and psychologists have extensively studied the personality of humans in order to understand the differences among individuals. According to Patrick and Leon [2], personality is a collective of perceptions, emotions, cognitions, motivations, and actions of an individual that interacts with various environmental situations. The definition showed that the personality is crucial effects that affect the livelihood of human and social system and plays indispensable role in human civilization. The evolution in technological landscapes that lead to the mass participation of users fuelled the study of personality in computing environments. There are three main domains particularly concerned with the integration of personality in computing. The three domains are 1) Automatic Personality Recognition (APR) that focusing in recognizing the true personality of an individual, 2) Automatic Personality Perception (APP) that concentrated on predicting the personality attributes and, 3) Automatic Personality Synthesis (APS) that generates artificial personalities through embodied Agents [18]. The methodology proposed in this study makes this paper more appropriate to be included in APP studies. As one of the Web 2.0 product, Facebook became one of the application that widely used by billions of users every day. The massive usage of Facebook has caused the creation of enormous amount of text data and indirectly raised the interests to study the roles of the words in identifying the behaviours of users. Due to the availability of "status messages" data, we started to study the personality of users by integrating the PEN Model as a personality framework. In the other hand, psychologists have developed different types of personality models or theories in order to illustrate the characteristics of human. Personally, we felt that the PEN Model matched with our research objectives that specifically attempting to identify the behaviours of criminals by exploiting the language cues that corresponds to their characteristics. Eysenck as a creator of PEN Model determined three main traits namely extraversion, neuroticism and psychoticism to represent the characteristics of humans in a general view [3]. The extraversions trait specifically reflects the positive affective states such as sociable, loving and enthusiastic while the neuroticism trait acts as a counter trait, by comprising the negative affective states such as hate, sadness and anxiety. The psychoticism trait is mainly associated with the characteristics of unempathetic, anti-social, and aggression behaviours. Numerous elements could act as an indicator to reveal the characteristics of humans. As one of the elements that indicates personality, researchers found that the linguistics patterns used by people to communicate either implicitly or explicitly signify their individual personalities. One of the earliest literature researchers, Sir Francis Galton hypothesized that natural language terms may represent actual personality differences among individuals [3]. Another researcher, Allport and Odbert claimed that almost 18,000 English terms could represent the different types of personality characters [3]. In 1990, Hofsteed have claimed that sentences and nouns probably have the connotations of personality [3]. This study incorporated one of the three-dimensional views [7] element called valence to measure the semantical connotation each of the candidate words. The valences measure the level of pleasantness and unpleasantness of each word. The idea to incorporate semantics in automatic personality detection rose due to the strong correlations existed between semantics and emotional words [10]. In spite of the fact of our principle that central on psychoticism, this preliminary study will expound the findings for all the traits from a questionnaire that executed to assemble the general perceptions of Malaysians on several words extracted from Facebook "status messages".

2. Related Work

Despite the fact that the investigations of psycholinguistics have begun long time back, the inventions of computer programs that can analyze the textual elements in more advance created a new vision to examined the personality of the people efficiently and effectively. From our literature studies, Philip Stone and his colleagues known as the first team that invented a computerized system called General Enquirer to analyze the textual contents related with behavioural sciences [27]. They also have been claim that the language used by human indirectly could expose the personality of the speaker. Overall, the strong relationship between language/writing styles and personalities that acted as a fingerprint could reveal the characteristics of the user. Another notable text analysis application called

Linguistic Inquiry and Word Count (LIWC) developed to measure the degree of using different categories of words by users across a wide array of texts data [13]. The development of text analysis applications, tools and techniques increases the effectivity and efficiency by reducing the time and efforts to examine the personality of online users when using a large amount of text data. Another study conducted by Argamon et al. known as an earlier work specifically focused on automatic personality detection task [16]. The work concentrated on classifying the extraversion and neuroticism traits using student essays from University of Texas by analyzing several linguistics features [16]. Oberlander and Nowson conducted a study to classify the personality of weblog authors using text by accessing the information from self-assessment report of the participants [16]. They applied several machine learning classification on Big 5 traits and found that some classification models perform well over the baseline. Since the studies of Argamon et al., scholars have been started to study the personality of users from various angles such as exploiting the linguistics features [20], classifying based on structure [21] and classification using various machine learning algorithms [21,23]. Most of the existing automatic personality detection studies employed the Big 5 [20, 23] personality model as a main framework to study the behaviours of users by integrating the self-assessment reports and the features extracted from their data. There are studies that utilized other models such as the Dark Triad Personality [19], Psychopathy [17] and many more. Similar to the aforementioned studies, most of the works in automatic personality detections applied various machine-learning algorithms by exploiting the information from self-assessment reports and linguistics markers. As far we are concern, there is no studies that exploited valences to measure sentiment of the words to detect the personality. Perhaps, the availability of information from self-assessments reports and linguistics markers substituted the role of the valence in identifying the personality of social networks users. On the other hand, we have found a work that used valence in predicting personality impressions from YouTube videos [14]. Moreover, the literature studies also revealed the valences also extensively studied in various disciplines. In 1999, Margaret and Peter have developed a textual corpus (ANEW) containing thousands of the affective rated by Introductory Psychology class students based on dimensional view attributes [7]. The results of the study presented by using mean and standard deviation for each dimensional view from the responses of male and female students. The study has become a benchmark for others in investigating the correlation of affective text in various fields. As one of our main reference, Nielsen has been developed a self-rated affective lexicons (AFINN) based on valences of thousands of the words collected from Twitter and evaluated the accuracy of classification with ANEW [8]. The AFINN vocabulary was select as our primary dictionary corpus for this study as the rating of the valences are indistinguishable, identical in terms of synonyms and provides admissible valence ratings by utilizing senses and folk psychology techniques. Another study by Vasa et al. studied the valence ratings of 81 emotional words among children aged from 9 to 11. They found that the female children have provided stronger valence ratings than male towards threat and positive words. Although, the result from the study is not having any effects to this study, the idea of using valence in appraising the emotional strengths is visible.

3. Methodology

Fig.1. depicts the methodology of this study to detect automatically the personality of Facebook users. First, the dataset was retrieve from the website of *Workshop on Computational Personality Recognition* (Shared Task) [11]. Second, the data cleaning process was carrying out to remove the unneeded textual elements such as URL, numbers and punctuations. In the same time, the spelling of the sentences corrected manually. Then, in order to remove the redundancy of words, we change all the words to lower cases and the Porter Stemmer Algorithm applied to the dataset in order to get the root words [12]. This step would reduce the words collection and assist the words short listing task. Thereafter, the Part of Speech (POS) Tagging applied to the cleaned dataset using Stanford POS Tagger. From the output of the POS tagger, we are more concentrated on the adjective words from the tagged dataset. The ability of the adjective words that have the significant effects in reflecting the opinion, emotions and thoughts could acted as a proximal cues to representing the personality of the users [6].

Then, we also extracted some of the nouns and verbs as possible cues that could mirror the psychological aspect of human being [29]. Next, we checked the semantics of all the words and only retain those words that having sentiment connotations. Then, we removed those words that already existed in our main reference corpus [8] so the questionnaire look compact, short and to minimize the corruptions in responses due to the fled halfway by respondents. Therefore, 51 English words and 16 interjections brought to the Likert scale questionnaire to obtain the general perceptions of Malaysians toward those words. In each page of the questionnaire, the instruction on how to choose the options had been attach with the example of valences extracted from [8]. Moreover, the scales of valence and its associations towards extraversion, neuroticism and psychoticism also have stated clearly, so that the respondents understands on how to answer each of the questions provided in the questionnaire. The scales ranged from 1 to 5 specifically will represent the sentiment strength of extraversion related word, -1 to -3 will represent the neuroticism related words while the scales of-4 and -5 will depict the psychoticism related words. Moreover, if the respondents felt that the particular words do not convey any sentiment strength, he/she can choose either "Not Sure" or "Not Applicable" option. In each of the pages in the questionnaire, the information on how the scales representing the personality traits had been state clearly. The information about gender, age, and and race of the respondents collected so that the demographical information will act as independent variable during analysis process. The result from the questionnaire will be elaborate in the next section.

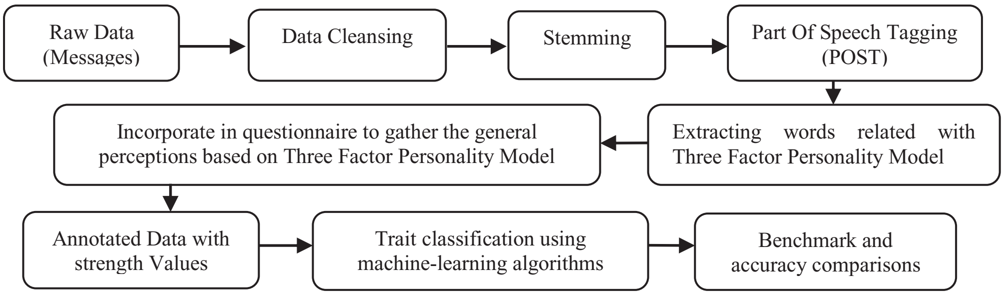

Fig. 1. The Methodology Diagram

After evaluating the validity and significance of categorizing the given words based on each trait, the selected words from this study and the sentiment words corpus from the work [8] will be use to annotate the Facebook "status messages". The annotated data will become an input to the classification tasks. In the automatic trait classification process, several machine-learning algorithms will be evaluate by using various parameters settings and linguistics features in order to find the best performing algorithms over the baseline by using the accuracy as evaluation metrics. Finally, the performances from each classification types will be discuss further to understand the characteristics of the data and its influences towards classification process.

4. Result and Analysis

Responses from 100 Malaysians randomly selected as sample for evaluation after remove the corrupted responses. The frequency of the demographics information are as follows ; gender (*male* : **0.50**; *female* : **0.50**), (age : *20-29 years* : **0.40**; *30-39 years* : **0.38**; *40-49 years* : **0.16**; *50 and above* : **0.06**), and race (*Malay* : **0.**44; *Chinese* : **0.**27; *Indian* : **0.25;** *Other* : **0.04).** The frequencies show that 78% of respondents ages in the range 20 to 40. The mean and standard deviation for the three demographics information are; gender (*mean* :**1.50**; *s.d* : **0.503**), age (*mean* : **2.88**; *s.d*: **0.891**), and race (*mean* : **1.89**; *s.d*: **0.920**). Table 1 showed the words and its frequencies of responses, mean, standard deviation, and valence of those words.

Table 1. The frequencies of responses, mean, standard deviation, and valence of neuroticism, psychoticism and extraversion words.

Neuroticism, Psychoticism and Extraversion based classified words											
Num	Word	N	Mean	S.D	V	**Num**	Word	N	Mean	S.D	V
1	Bark	85	2.26	1.416	-2	**35**	Arghhh	72	2.71	1.144	-3
2	Cringe	76	1.78	0.988	-2	**36**	Eee	63	2.68	1.060	-3
3	Desolate	72	1.86	0.939	-2	**37**	Eek	69	2.78	1.199	-3
4	Booty	64	3.00	0.943	-3	**38**	Ehhh	67	2.97	1.073	-3
5	Cum	69	3.99	1.064	-4	**39**	Erghhh	63	3.14	1.176	-3
6	Creepy	74	2.35	1.152	-2	**40**	Garggg	74	2.89	1.105	-3
7	Cranky	70	2.27	1.102	-2	**41**	Grrr	66	3.08	1.100	-3
8	Buttseck	89	3.09	1.285	-3	**42**	Ughhh	64	2.83	1.292	-3
9	Arse	74	2.97	1.060	-3	**43**	XXX	68	3.79	1.073	-4
10	Discomfort	67	2.16	1.053	-2	**44**	Fie	65	2.89	1.017	-3
11	Debauchery	77	2.78	1.373	-3	**45**	Crave	86	2.42	1.111	2
12	Delude	74	2.68	1.251	-3	**46**	Credible	89	2.72	1.128	3
13	Devious	80	2.69	1.186	-3	**47**	Dazzle	87	2.78	1.115	3
14	Disgruntled	77	2.57	1.152	-3	**48**	Decent	89	2.91	1.184	3
15	Feral	67	2.46	1.105	-2	**49**	Diligent	85	3.12	1.276	3
16	Filthy	80	2.71	1.046	-2	**50**	Endure	92	3.09	1.298	3
17	Gruesome	75	2.57	0.975	-3	**51**	Gorgeous	91	3.41	1.374	3
18	Goddamnit	81	4.30	1.112	-4	**52**	Gratitude	96	3.18	1.384	3
19	Hatred	77	2.75	1.114	-3	**53**	Groovy	93	3.16	1.296	3
20	Grouch	68	2.54	1.028	-3	**54**	Kickass	88	3.38	1.216	3
21	Grudge	75	2.16	1.139	-2	**55**	LMBO	80	3.15	1.181	3
22	Irk	73	2.12	0.832	-2	**56**	Neat	90	3.23	1.333	3
23	Lazy	69	2.28	0.856	-2	**57**	Persevere	88	3.01	1.246	3
24	Irksome	73	2.44	0.866	-2	**58**	Squee	86	3.14	1.160	3
25	Lousy	74	2.23	0.987	-2	**59**	Stellar	91	3.21	1.269	3
26	Pester	74	2.35	1.026	-2	**60**	Unafraid	86	3.12	1.162	3
27	Maniac	78	2.79	1.144	-3	**61**	Woot	90	3.27	1.178	3
28	Mofo	85	4.53	0.946	-5	**62**	Ahhh	87	3.01	1.225	3
29	Penis	82	4.05	0.915	-4	**63**	Eww	75	2.69	1.185	3
30	Shudder	74	2.84	1.135	-3	**64**	Hahaha	87	2.93	1.189	3
31	Slap	70	2.71	0.995	-3	**65**	Ohhh	92	2.90	1.187	3
32	STFO	76	3.16	1.337	-3	**66**	Wahhh	84	2.90	1.219	3
33	Unkind	77	2.87	1.281	-3	**67**	Wahoo	94	3.24	1.224	3
34	Wack	79	2.95	1.348	-3						

4.1.1 Reliability of ratings by respondents

In order to examine the reliability of scales rating by the respondents, the internal consistency measuring testing conducted by applying Cronbach's alpha coefficients. The Cronbach's Alpha Coefficient for extraversion related words are 0.951, neuroticism words are 0.937, and psychoticism words are 0.855. The responses for the individual word also strongly correlated with the total responses to each of the words. The correlations for the rating patterns of negative words ranged from 0.63 to 0.89 whereas for the positive words ranged from 0.75 to 0.96. The high values of Cronbach's alpha coefficients and rating patterns by respondents against individual words it delineate higher degrees of reliabilities and confidences in responses and categorization of 67 words under each of the PEN model traits.

4.1.2 Comparison of significant between words related with traits and demographics data.

The multivariate analysis has executed in order to determine the rating significances between words under each of the traits against age, gender, and race. By using the Wilk's lambda approach, we found the significant multivariate effect for psychoticism words for age, λ=(0.487), F(15,127) = 2.532, $p < 0.05$, $n^2 = 0.17$ and gender λ=(0.453), F(17,112) = 2.453, $p < 0.05$, n^2=0.15. Moreover, the words from extraversion trait shows significant effects against gender, λ=(0.587), F(17,42) = 1.739, $p<0.05$, n^2=0.41.

Then, we found that neuroticism words strongly signify with the age, λ=(0.128), $F(27,88) = 3.339$, $p<0.05$, $n^2 = 0.49$. The results showed that gender and age are important factors that cause divergences in perceptions towards the connotations of the words and influence the perceptional differences among human being. This could verify by looking at the LIWC application where the gender and age information become subjects of consideration during the text analysis task [13]. Although, there is no significant found between words categorization and races, it is does not means the diversity in races have no significant towards words categorization. We believed that there are mild effects between words perceptions and races especially to the words that brought higher negative connotations.

Next, we will discuss the distribution of mean values from the rating of words for all the traits by using age and gender as independent variables. The mean distribution for extraversion words showed that the male respondents appraised to the higher positive valences contrast to the evaluation by female that more centred in middle positive values. This result contradicts with the results from the study conducted by Vasa et al [9] where the female children rated higher to the positive words. Perhaps, the contradictions occurred due to the differences in perceptions between children and adults or the nationality effects cause such divergences. In the same time, the mean value also revealed that the male aged between 40 - 49 years is less sensitive towards the semantical connotations of positive words. Yet, although our analysis does not found any significant effects between neuroticism words and gender, we are still optimistic that there is moderate impact between neuroticism words and gender. By analyzing the mean distribution, we found that female appraisals for the neuroticism words are higher as 5.6% than male. The pattern of the mean showed that female participants are more sensitive to neuroticism words than extraversion related words. The mean differences from age group showed that participants with the age category of 20 - 29 years and 30 - 39 years rated the neuroticism words as much as 5.7% to 12.6% higher than other categories of age. Fig.2 showed the mean distribution of rating of psychoticism words. Even though, the bar graph showed that rating patterns by both male and female from all categories of ages centred on higher negative valences, further analysis showed that female's average ratings 3.9% greater than male. In addition, the differences in terms of mean values against age showed that the participants aged between 30 - 39 years and 40 - 49 years rated the psychoticism words 7% to 10.7% higher than other category of ages. Laterally, the rating patterns reflects the perceptions of respondents towards those words as highly negative words and indirectly provide evidence that those words might easily causes hurt to the people feelings and could cause conflicts and controversy more easily. By applying the Brunswikian Lens model [28], the words categorized under psychoticism words is perfectly reflect the characteristics stated in psychoticism. For instance, the semantics of the words of *mofo* and *goddamnit* obviously reflect the aggression and anti-social behaviours and could be use as cues to detect the psychoticism traits among Facebook users.

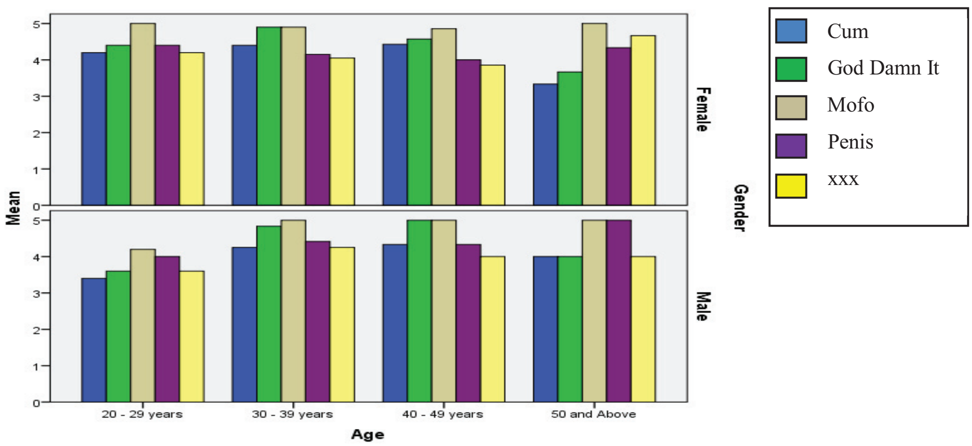

Fig. 2. The distributions of mean values of psychoticism words across age and gender.

5. Discussion

The idea of the APP in social networks that exploits the perceptions of others towards certain people based on their usage of words and language styles is one of the challenging tasks. The perceptions driven technique could solve several issues raised due to unavailability of conclusive self-assessment reports. For instance, the techniques could hinder the divergences in emotional perceptions occurred because of the nationality effects and biases causes by self-assessment reports. Nevertheless, the result from this study have described distinct perceptions in terms of demographical information, we are just using the majority perceptions without depending on any specifics demographical information. Then, the strong internal consistency of measuring the rating patterns perfectly reflect the interrelationships existed among the psychoticism words. Moreover, the higher mean values from the valence rating either from male or female respondents reflect the stronger negative sentiment of psychoticism words. The Cronbach's alpha coefficients, Wilk's significant, and mean values for the psychoticism related words strengthen our strategies to make those words as cues in the our main study that focus on classifying the personality of Facebook users by applying machine learning approaches. In the other hand, the semantics of the psychoticism words that mirror the negative sentiments perfectly reflecting the definitions stated in the language of Anti-Social Behaviours. According to Fitzgerald, the language expression of anti-social behaviours is deeply value laden, implying purposeful negative action and harmful to others [24]. In here, the negative actions could perceive as the actions that used higher negative words that might offence others feeling. Such words are very common in most of the online bullies and harassment cases. This could be proving by observing the techniques employ by scholars to detect cyber bullies [25] and harassment detection studies [26]. The anti-social behaviours and cyber bullies [1] suspect behaviours also match perfectly with the characteristics of psychoticism. Therefore, we believe that automatic psychoticism trait detections among online users could reduce the number of the cyber bullies and cyber harassment cases. Moreover, our word collection also could be use by other studies from sentiment analysis, personality detection, authorship detection, bully detection, spam filtering and et-cetera. Furthermore, the tens of the word collections could enrich the current lexical corpus.

6. Conclusion

This study practically present the approach to identify the personality of users based on general perceptions from Malaysians perspectives. The tens of the words extracted from Facebook status messages became subject of the study to gather the majority perceptions based on dimensional view called valence. In general, male participants rated to the extraversion words much higher than female and as contrast, the female respondents rated to neuroticism words much higher than male participants did. Nonetheless, the female rated a little higher to the psychoticism words than male, we believed that there are similarities between male and female in rating of psychoticism words where almost, all the participants chose the higher negative scales. The similarities indirectly showed the general perceptions of Malaysia respondents were all of their perceptions collectively homogeneous. The homogenous perspective with the result from analysis using Cronbach's alpha coefficients, Wilk's significant and mean values strengthen our ideas to use the higher negative words as cues to detect the psychoticism trait among Facebook users.

Acknowledgements

We would like to thank all of our survey respondents and University Sains Malaysia for providing support to this research work.

References

[1] Osaman Tolga Aricak. (2009). Psychiatric Symptomatology as a Predictor of Cyberbullying among University Students. *Eurasian Journal of Educational Research*, 34, pp. 167-184,Winter 2009.
[2] Ellis. (2008). The Study of Personality. Accessed From http://www.sagepub.com/upm-data/23238_Chapter_1.pdf.
[3] Gerald Matthews, Ian Deary and Martha C. (2003). Personality Traits (2nd edition).*Cambridge University Press*.
[4] Grainne Kirwan and Andrew Power. (2011). The Psychology of Cyber Crime: Concepts and Principles. ISR.
[5] Mohammad Rahim Kamaluddin, Nadiah Syariani Md Shariff, Azizah Othman, Khaidzir Hj Ismail and Geshina Ayu Mat Saat. (2015). Linking psychological traits with Criminal behavior: a review. *Linking Pychological Traits With Criminal Behaviour: A Review ASEAN Journal of Psychiatry,* Vol. 16 (2), July - December 2015: XX-XX.
[6] F. Mairesse and et.al. (2007). Using Linguistic Cues for the Automatic Recognition of Personality in Conversation and Text. *Journal of Artificial Intelligence Research* 30 (2007) 457-500.
[7] Bradley, M.M., & Lang, P.J. (1999). Affective norms for English words (ANEW): Instruction manual and affective ratings. Technical Report C-1, *The Center for Research in Psychophysiology*, University of Florida.
[8] Finn Arup Nielsen. (2011). "A new ANEW: Evaluation of a word list for sentiment analysis in microblogs", *Proceedings of the ESWC2011 Workshop on 'Making Sense of Microposts': Big things come in small packages 718 in CEUR Workshop Proceedings* : 93-98. 2011 May.
[9] R.A. Vasa et al. (2004). Valence ratings of emotional and non-emotional words in children. *Personality and Individual Differences*. 41. (2006) 1169–1180.
[10] Oberlander, J., and Nowson, S. (2006). Whose thumb is it anyway? classifying author personality from weblog text. *In Proceedings of the 44th Annual Meeting of the Association for Computational Linguistics (ACL).*
[11] Kosinski, M., Stillwell D.J., and Graepel T. (2013). Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the National Academy of Sciences (PNAS)*.
[12] Porter, M.F. (2001), Snowball: A Language for Stemming Algorithms. Available at: http://www.snowball.tartarus.org/texts/introduction.html.
[13] Yla R. Tausczik and James W. Pennebaker. (2010). The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods. *Journal of Language and Social Psychology* 29(1) 24–54.
[14] G. Sonja and K. Koroveshovski. (2014). The Impact of Affective Verbal Content on Predicting Personality Impressions in YouTube Videos. *Workshop on Computational Personality Recognition 2014*.
[15] Philip J. Stone and Earl B. Hunt. (1963). A computer approach to content analysis: Studies using the General Inquirer System. *In proceeding Spring Joint Computer Conference*.
[16] Argamon, S., Dhawle, S., Koppel, M., and Pennebaker, J. (2005). Lexical predictors of personality type.*In Proceedings of the Joint Annual Meeting of the Interface and the Classification Society of North America.*
[17] Randall Wald, Taghi Khoshgoftaar, and Amri Napolitano. (2012). Using Twitter Content to Predict Psychopathy. *International Conference on Machine Learning and Applications*.
[18] Alessandro Vinciarelli and Gelareh Mohammadi. (2013). A Survey of Personality Computing. IEEE Transactions On Affective Computing.
[19] Danilo Garcia and Sverker Sikstrom. (2013). The dark side of Facebook: Semantic representations of status updates predict the Dark Triad of personality. *Personality and Individual Differences*. 67. (2014) 92–96.
[20] Golbeck, J., Robles, C., and Turner, K. (2011). Predicting Personality with Social Media. *In Proc. of the 2011 annual conference extended abstracts on Human factors in computing systems*. 253–262.
[21] Francisco Iacobelli and Aron Culotta. (2013). Too Neurotic, Not Too Friendly: Structured Personality Classification on Textual Data. AAAI Technical Report WS-13-01 Computational Personality Recognition (Shared Task).
[22] Grainne Kirwan and Andrew Power. (2011). The Psychology of Cyber Crime: Concepts and Principles. ISR.
[23] Golbeck, J., Robles, C., Edmondson, M., and Turner, K.2011(b). Predicting Personality from Twitter. In Proc. of International Conference on Social Computing.149–156.
[24] Fitzgerald, M. (2011). Submission to the Department of Human Services on behalf of Public Housing Tenants in relation to Human Rights concerns raised by the Anti-Social Behavior Pilot. Fitzroy Legal Service.
[25] April Kontostathis, Kelly Reynolds, Andy Garron, and Lynne Edwards. (2013). Detecting cyberbullying: query terms and techniques. *WebSci '13 Proceedings of the 5th Annual ACM Web Science Conference*.
[26] Dawei Yin, ZhenzhenXue, Liangjie Hong, Brian D. Davison, April Kontostathis, and Lynne Edwards. (2009). Detection of Harassment on Web 2.0. *In Proceedings of the Content Analysis in the WEB 2.0 (CAW2.0) Workshop at WWW 2009.*
[27] Philip J. Stone and Earl B. Hunt. (1963). A computer approach to content analysis: Studies using the General Inquirer System. In proceeding Spring Joint Computer Conference.
[28] Wolf, B. (2005). Brunswik's original lens model.
[29] Yla R. Tausczik and James W. Pennebaker. (2010). The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods. Journal of Language and Social Psychology. Vol: 29(1) 24–54.