Value of Expressions behind the Letter Capitalization in Product Reviews

Irina Pak
School of Science and Technology,
Sunway University, Bandar Sunway,
Selangor, Malaysia
+6-0132101373
irina.p@imail.sunway.edu.my

Phoey Lee Teh
School of Science and Technology,
Sunway University, Bandar Sunway,
Selangor, Malaysia
+6-0166182878
phoeyleet@sunway.edu.my

ABSTRACT
Product reviews from consumers are the source of opinions and expressions about purchased items or services. Thus, it is essential to understand the true meaning behind text reviews. One of the ways is to analyze sentiments, expressions and emotions behind the text. However, there are different styles of writing used in the text. One of widely used in the text is letter capitalization. It is commonly used to strengthen an expression or louder tone within the text. This paper explores the value of expression behind letter capitalization in product reviews. We compared fully capitalized text, text with one capitalized words and text without capitalization through the readers' perspective by asking them to rate the text based on Likert scale. Furthermore, we tested two samples of text with and without capitalization on 27 available online sentiment tools. Testing was done in order to check how current sentiment tools treat letter capitalization in their sentiment score. Results show that of letter capitalization is able to enforce the different level of expression. If the nature of the review is positive, the capitalization makes it more positive. Similar for the negative reviews, the capitalization tends to increase negativity.

CCS Concepts
• **Information systems~Social networks** • **Information systems~Online shopping**

Keywords
Capital letters; Uppercase; Text analysis; Opinion mining; Sentiment analysis

1. INTRODUCTION
Identifying expressions or emotions from the text is referred as sentiment analysis. There are several reasons to analyze those expressions. Comments in social media, feedbacks, and discussion on forums often express the opinion and experience of products

purchased and services received. Tang et al. [13] pinpointed that there is a need to analyze it due to the rapidly growing number of social media texts used in the market for opinion mining. With the freedom of speaking online as according to Moreo et al. [7] , with or without anonymously referred, the expression is normally enforced by the writer using various ways of writing. Derks et al. [4] have also confirmed that people tend to express emotions in online communication, and the frequency is higher compare to face-to-face communication. Hence, the consumers commonly make the social media texts, such as online reviews as their referral before their next purchase. Thus, analyzing expressions and emotions within the text is essential, and it has made many researchers keep trying to improve their sentiment tools with the different methods.

The current studies in sentiment analysis should endorse the need for sentiment tools to incorporate the value of the expression. In this case, we are discussing the expression that embedded in the social media text. There is another study that tends to enforce the need to detect emotion, but not only to the aspect of topic-related or aspect extraction [9]. Another study also looks at how the emoticons have reversed the meaning of comments Teh et al. [14]. Moreover, Sanchezrada and Iglesias [12] also highlighted the importance of the emotions expressed within the text. They stated that emotion detection is useful in brand monitoring. In another case, punctuation can use as a form of expression within the text. For instance, it has been identified that there is a need for exclamation marks to be included in sentiment tools for analysis because it has a value that enforced the expressions within the text [15]. In many reviews and comments, people tend to use different styles of writing to express their emotions within the text. In this study, the measurement of those emotions and expressions is referred as value expression.

In this study, we focus specifically on the utilization of letter capitalization or upper case in product reviews. This study presents comparison on the impact of expressed value from following capitalization formats: 1) text without letter capitalization, 2) text with one capitalized word and 3) fully capitalized text. The contribution of this study is to identify value expression behind the letter capitalization in order to find more accurate and detailed technique to analyse the social media text.

2. MOTIVATION AND RELATED WORK
Capitalization of the word(s) can strengthen the expression in the text. For instance, capitalized word used in the middle of the sentence or capitalized sentence can indicate the stress [3]. Another

example of capitalization used for emotional expression is addressed by Anil et al. [1]. According to them, one of the factors to detect emotions is referring to the user style of typing including the usage of letter capitalization. They stated that capitalization is applied more often when the expression is in the category of mood, which is 80% of the cases. Similar to the study of Byron [2] which agreed that capitalization could be used to express emotions in text. In fact, usage of capitalization indicates how intense the emotions are [2]. Moreover, Ledbetter et al. [6] stated that fully capitalized text has interpreted as shouting and unwelcome emotion. However, whether the usage of capitalization is used to express a happy mood or unwelcome emotions, it has impact on the reader opinions when

brackets, underscores, curly braces, repeating exclamation points, and underlines also played a role in determining the sender message's intention. The nonverbal meaning of a social media text has seemed far more important that has it sometimes can even reverse the overall meaning of comments, which could only be identified by a human being.

3. METHODOLOGY
3.1 Data collection
We have collected social media texts from online reviews manually. There is a total of 1014 reviews based on 10 different
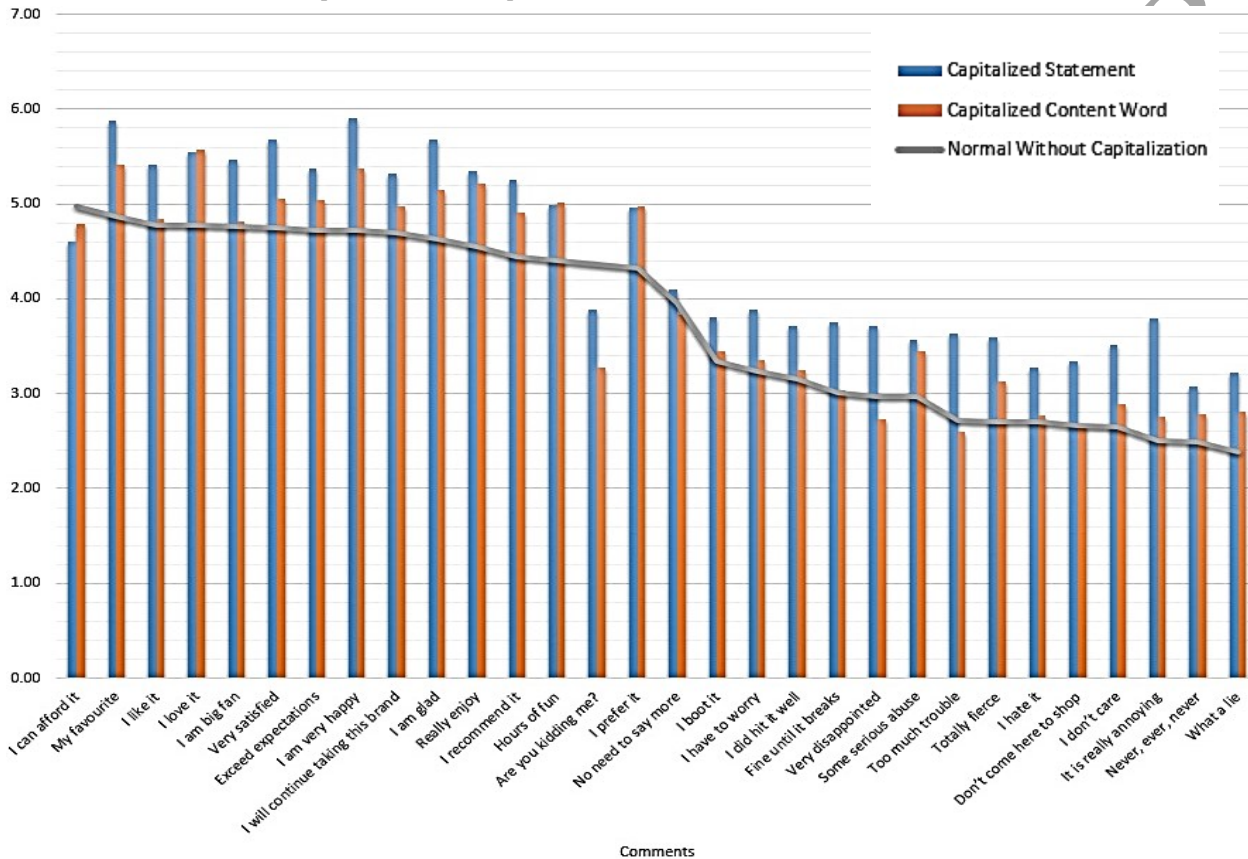

Figure 1 The Impact of Capitalized Social Media Text towards Purchase Likelihood

a comment or product review is read online.

Until today, the textual modifiers of the social media text by the writer is assumed to be received from the reader or to show the emotional state of the writer. Which in vice versa, sentiment tools that are developed to process the meaning of words in opinion mining thus starting to assist human in scraping the semantic/sentiment meaning of all social media texts from social media sites. Obviously, the emotional state of the writer is often ignored in many cases, with the advancement of tools that automatically remove or repairing all the grammatical error in the text, and returning the word to low case, instead of checking the meaning of words which are capitalized and rate them accordingly.

Further, the use of capital letters can underline meaning and characterize the sign of sarcasm Ledbetter and Larson [5]. That is even harder for a sentiment tool to process. In many cases, it is ignored. Similarly, Riordan and Kreuz [11] have also discussed the possibility of capitalization to specify a different meaning in the text. Besides capitalization, the usage of asterisks, braces, angled

types of products, which includes the categories of beauty and health, camera, computer, consumer electronics, fashion, home appliance, jewellery, watches, mobiles, tablets, sports materials, toys and kids reviews. The aim of collecting these various categories of reviews and not using any corpus is because we need these written reviews and comments to be raw and unprocessed, this is to keep its originality of how the comments were originally written in the first place by people. We summarize all the reviews based on various styles of the usage of a capital letter. Moreover, we have regrouped all these varieties into three groups. The first group includes 70 reviews that have all texts without any capitalization (e.g. including the first character of the first words that has not capitalized), we assume this type of social media texts could commonly be writen from a mobile phone. The second group has a total of 679 reviews with at least one word in capitalization, often the content word that they usually use to enforce with a capital letter to strengthen their emotions (e.g. i LOVE this product). And the third group has a total of 9 statements that comes with the whole statement in capital letters. Lastly, the fourth group with 256 social

media texts that have at least one capital letter within the sentence, usually the first word. With all these 1014 reviews, the third group has the smaller number in percentage, but the second group has the higher proportion.

3.2 Survey Set Up

With this set of collections from social media texts above, we designed a questionnaire to explore if the usage of letter capitalization has an impact on the reader's opinion. Using the sample of social media texts, we collect the opinion from the users through a questionnaire. On their likelihood of purchasing the products after reading the same reviews written in capitalization formats. Which also helps to understand the impact of how the usage of capitalization from writer influences the reader opinion. First, we selected top 30 highest frequently used emotional words (15 positives and 15 negatives), using semantic tools by Rayson [10]. Then, we extracted the statement with each emotional word concordance and included the statement in our questionnaire.

Table 1. Positive and negative reviews for the survey

Positive	Negative
i love it	some serious abuse
i like it	very disappointed
i am very happy	i don't care
i am glad	i did hit it well
i am big fan	i hate it
my favourite	it is really annoying
hours of fun	i boot it
very satisfied	too much trouble
i prefer it	totally fierce
really enjoy	i have to worry
i recommend it	i can afford it
exceed expectations	what a lie
i will continue taking this brand	don't come here to shop
are you kidding me?	fine until it breaks
no need to say more	never, ever, never

The list of selected statements is shown in Table 1. Then, we simulated the statement with three capitalization formats. Five hundred respondents were requested to rate after they have read the text. Likelihood includes 7-points scale from 1-strongly dislike till 7-strongly like.

3.3 Sentiment Tools Set Up

Sentiment tools' examination was accomplished in order to explore how the currently available sentiment tools handle the letter capitalization in their sentiment score.

Table 2 Sentiment tools used in this study

No[1]	Name of the tool	Web Source
1.	Selasdia Intelligent Sales Assistant	http://www.aiaioo.com:8080/annotator-0.1/automation/demoView/1
2.	Sentaero	http://www.sentaero.com/textsearch.php
3.	Meaning cloud	http://www.meaningcloud.com/demo
4.	TheySay	http://apidemo.theysay.io/
5.	Repustate	https://www.repustate.com/api-demo/
6.	Text sentiment analyzer	http://werfamous.com/sentimentanalyzer

[1] Referring numbers of the tools are the same as in Table 4

7.	MIOPIA Supervised Model	http://miopia.grupolys.org/demo
8.	SentiStrength	http://sentistrength.wlv.ac.uk/
9.	Python NLTK Demos for Natural Language Text Processing	http://text-processing.com/demo/
10.	LIWC	http://liwc.wpengine.com/
11.	Sentiment Analyzer	http://www.danielsoper.com/sentimentanalysis/#
12.	Sentiment Analysis Opinion mining	http://text2data.org/Demo
13.	Pattern Sentiment Analysis	http://textanalysisonline.com/pattern-sentiment-analysis
14.	Sentiment Vivekn[8]	http://sentiment.vivekn.com/
15.	Alchemy Language Document Sentiment	https://alchemy-language-demo.mybluemix.net/
16.	Alchemy Language Targeted Emotion	
17.	Intellexer	http://demo.intellexer.com/
18.	ParallelDots	http://www.paralleldots.com/sentiment-analysis
19.	DepecheMood	http://www.depechemood.eu/DepecheMood.html
20.	Twinword	https://www.twinword.com/api/sentiment-analysis.php
21.	uClassify	https://www.uclassify.com/browse/uclassify/sentiment?input=Text
22.	Tone Analyzer	https://tone-analyzer-demo.mybluemix.net/
23.	Pythia Semantic Features	
24.	Pythia Term n-grams	
25.	Pythia Character n-grams	http://omiotis.hua.gr/pythia/#
26.	Pythia All n-grams	
27.	Pythia All Features	

List of the tools is presented in Table 2. This study limited to 27 online available sentiment tools. We selected one review comment from the collected 1041 selected ones for sentiment tools' testing.. The selected comment is "*Very good but too expensive. Excellent product but too pricey. I bought it to my son who uses it for graphic design. He loves this product but I would not buy it for me because it is too expensive*". The second selected comment was the same text but fully capitalized. So, testing includes comparing scores between low case and fully capitalized text. That helps to check if there is any difference in tools' score when the letter capitalization is used. The comparison of the score is made across the same tool and not between the different tools. Therefore, normalization of the score is not needed for this paper analysis.

4. RESULTS

4.1 Survey Results

Figure 1 marks overall mean for 500 respondents' rating based on the usage of capital letters in various scenarios. From most like to most dislike, we observe from Figure 1 that most texts with a higher like value are mainly positive in polarity. Such as the word "I am

very happy", "my favourite" and "I am glad". And looking from the less likely to purchase a product, which is below 3, most comments are negative polarity, such as the word "Never, ever and never", "what a lie" and "I hate it". From Figure 1, the impact for the positive texts has shown a clear Likelihood in opinion, mostly above 3 of the Likert scale. We observed that both capital letters used for the whole statement content words in capital letters had its impacts toward the reader. Respondents had answered with a higher likelihood to purchase after reading those texts in capital letters. In addition, when it comes to the negative comments, surprisingly, it showed a slightly different opinion on texts that appear in negative polarity. The result indicates that there is contradictory even the statement is negative in polarity, and with a capital letter in use.

Interesting that letter capitalization used for negative statement intrigues a higher likelihood. Even greater comparing to text appeared in it low case letters.

Table 3 Correlation statists

Tested variations	Pearson Correlation	Sig. (2-tailed)
NoCapitalization - FullyCapitalized	0.427**	0.000
NoCapitalization Positive - FullyCapitalized Positive	0.431**	0.057
NoCapitalization Negative – FullyCapitalized Negative	0.646**	0.000
NoCapitalization - OneWordCapital	- 0.133**	0.003

The negative 0.133 coefficient demonstrates a connection between these two variables in the same opposite way that positive correlation coefficient does, with the relative strengths of likelihood are the same of 0.133. In others words, when a written comment is in capital letters, it has certainly impact and influence on the reader, and it is in 2-tailed with a significant value of 0.003.

4.2 Sentiment Tools' Results
Bellow table presents results from sentiment tools testing.

Table 4 Sentiment tools' results

No Tools	Not capitalized	Fully capitalized
1.	P[2]	P
2.	6.67	6.67
3.	NU[3] 92%	NU 92%
4.	P 0.707 NU 0.017 N[4] 0.277	P 0.583 NU 0.014 N 0.403
5.	-0.969013935029	-0.201760974677
6.	18%	18%
7.	Overall -8.7	Overall 29
8.	P 4 N -2	P 4 N -2
9.	Overall P P 0.7 N 0.3	Overall P P 0.7 N 0.3
10.	P 7.7	P 7.7
11.	N -100	N -100
12.	P +0.201	P +0.922
13.	P 0.182	P 0.182

NoCapitalization Positive – OneWordCapital Positive	0.110*	0.014
NoCapitalization Negative – OneWordCapital Negative	0.291**	0.000
OneWordCapital - OneWordCapital	0.605**	0.000

Furthermore, we continued to explore the significance of our study among all these scenarios. Table 3 shows all the results processed with the computational values organized in the way that is shown in the correlations (first column of Table 3). "NoCapitalization" is referred to a text without any capitalization. We computed by summarizing up all the 500 respondents likelihood results for both positive and negative comments and separate them accordingly to the range of positive to negative, drawing and organizing similar to the statement from Figure 1.

From the results, they are all significant. The highest correlations that highly correlate with each other is when we correlate the not capitalized negative text with the text that contains all capital letters, which is 0.65 with a 2-tailed significant value of 0.000. Other results overall have shown a significant impact of all of the not capitalized social media text (both in positive and negative polarity) and have a negative correlation with all the social media text that contain a particular word in capital letters (for both in positive and negative polarity).

14.	P 92.9176	P 98.9723
15.	N -0.181829	N -0.181829
16.	Anger 0.014415 Disgust 0.02732 Fear 0.03122 Joy 0.712898 Sadness 0.203233	Anger 0.122237 Disgust 0.029684 Fear 0.073817 Joy 0.383307 Sadness 0.50946
17.	NU 33.3% P 66.7%	NU 33.3% P 66.7%
18.	N	N
19.	Amused 0.949 Angry 0.119 Annoyed 0.507 Don't Care 0.59 Happy 0.602 Inspired 1 Sad 0.037	Amused 0.999 Angry 0.191 Annoyed 0.417 Don't Care 0.822 Happy 0.58 Inspired 1 Sad 0.07
20.	P 0.13296846986111	P 0.13296846986111
21.	N 85% P 15%	N 85% P 15%
22.	Joy 0.71	Sadness 0.51
23.	First two sentences are P Last two are N	First two sentences are P Last two are N
24.	N	N
25.	N	N
26.	N	N
27.	N	N

---

[2] P is referred to Positive in Table 4

[3] NU is referred to Neutral in Table 4

[4] N is referred to Negative in Table 4

According to the Table 4, overall results suggest that only 30% of the tools, which is 8 out of 27 showed different scores for the fully capitalized score. The content of the tested sample was overall positive, but with a hint of complaining about the high price. So with that in mind, we can evaluate the tools' scoring. Some of the tools give the only polarity as a result. For instance, Selasdia Intelligent Sales Assistant rated sample as positive, but tools Parallel Dots, Pythia Term n-grams, Pythia Character n-grams, Pythia All n-grams and Pythia All Features agreed that tested text is negative. Furthermore, Pythia Semantic Features tool gave more detailed polarity result, where first two sentences were rated as positive and the rest as negative. However, those tools gave the same polarity to text with and without capitalization.

Another tools gave the result as polarity and score. But results remain the same for the text with and without capitalization. For instance, Sentaero, Meaning cloud, Text sentiment analyser, SentiStrength, Python NLTK Demos for Natural Language Text Processing, LIWC, Sentiment Analyzer gave the same score and polarity to the fully capitalized text and low case text.

However, there are tools which gave different sentiment score with polarity to two capitalization formats. For example, tool TheySay shows that letter capitalization can make the text less positive and neutral, but more negative. Also, Tone Analyzer suggests that capitalization makes the text more negative since the results changed from joy to sadness. Sentiment tools MIOPIA Supervised Model, Sentiment Vivekn disagree with previous tools by suggesting that fully capitalized text is more positive than low case text. Two sentiment tools like Alchemy Language and Targeted Emotion DepecheMood gave detailed results with mood identifiers and scoring. Therefore, both tools agree that letter capitalization made the text more negative since the score for negative identifiers like "Anger", "Disgust", "Sadness", "Fear", "Sad", "Angry" and "Don't Care" increased. Moreover, scores for positive identifiers like "Joy" and "Happy" decreased.

Overall, we can see that most of the tested sentiment tools need to incorporate letter capitalization in their sentiment score. However, sentiment tools which give only polarity result cannot be judged equally in this content.

5. DISCUSSION

For further investigation, we extracted some examples of comments 1014 reviews. We observed that the utilization of capitalization is not only used in demonstrating user's expression, but also enforced the concentration of meaning of the particular capitalized word. It is also used in short formal words. There are mainly two groups where letter capitalization. First is the use of capitalization to enforce the emotions, especially to strengthen the intensity of the words.

S1: *its not a good choice...I think its PERFECT... :) Have Fun.*
S2: *I used this iron today and LOVE it - it takes out the wrinkles* in cotton - something many of the other irons don't do well.
S3: *This iron was not worth the money AT ALL. I purchased the iron about a month ago, and already it does live up to its name or functionality.*
S4: *Samsung makes the BEST TV's there are.*

In S1 and S2, capitalization is used intentionally to enforce the meaning of words "perfect" and "love". The impact of using a low case format for "love" and the capitalization in the statement has certainly conveyed a less intended expression. For S3 and S4, capitalized phrases have now enforced the attention to the reader to the strength of the level of expression, allowing our imagination of its enforcement in writing.

S5: *DON'T BUY IT ON AMAZON! The one you can buy on Amazon, Ebay... IS NOT an authorised Magic Bullet. I bought it, and it brakes after 2 months...*
S6: *Best scull caps sold at AX. NEVER FAILS. Love it!*
S7: *I Absolutely LOVE this Coach Leather Turnlock Tote Handbag!!!! I*
WANT TO OWN IT SO VERY BAD!!! It's a Terrific Giveaway! Thanks, Michele...
S8 :*I LOVE THESE SPEAKERS!!!!!, these speakers are Wonderful!!! Great sound quality and they are loud but only if u want them to be...*

Other than this, for S5 and S6, capital letters have been understood at first glance of all the capital letter social media text, highlighting the meaning of the whole statement. Thus, we can assume this as the main enforcement of overall statement in our opinion mining method. By reading only capitalized phrases, it is enough to understand the user satisfaction or dissatisfaction level about the product. For S6, despite the long statement with 10 words, the character of the capitalized word score double in a very optimistic and confident expression. S7 and S8, as understood by a human, has definitely presented the positive satisfaction towards the products.

6. CONCLUSION

In conclusion, the use of letter capitalization is clearly used to enforce the expression and underline the meaning of texts. It has significantly shown an increased level of expression comparing to the text without letter capitalization. If the review is positive in polarity as a whole, then the capitalization makes the overall meaning more positive. It indicates that capitalization used for positive texts able to signal a higher positivism. Similar to negative text, when the letter capitalization is used for negative text, text has a higher level of negativity. Results from questionnaire have confirmed that the usage of capitalization has its level of impact and influence, comparing to low case letters. However, there can be several other ways to explore the usage of letter capitalisation more detailed, for example creating a framework that's able to compare value expression of upper and low case text. That is a great opportunity for the future study.

Another conclusion can be made, based on the sentiment tools result; we can say that improvement of sentiment analysis is needed. The improvements should include the consideration of letter capitalization in their sentiment scoring. That leads to the enhancement of current sentiment analysis tools towards exploring different formats and styles of writings.

7. ACKNOWLEDGMENTS

8. REFERENCES

[1] Anil KKM, Kiran BR, Shreyas BR, Victor SJ (2015) A Multimodal Approach To Detect User ' s Emotion. In: 4th International Conference on Eco-friendly Computing and Communication Systems. Elsevier Masson SAS, Kurukshetra, pp 296–303

[2] Byron K (2008) Carrying too heavy a load? The communication and miscommunication of emotion by email.

Acad Manag Rev 33:309–327

[3] Carey J (1980) Paralanguage in computer mediated communication. ACL '80 Proc 18th Annu Meet Assoc Comput Linguist 67–69

[4] Derks D, Fischer AH, Bos a. ER (2008) The role of emotion in computer-mediated communication: A review. Comput Human Behav 24:766–785. doi: 10.1016/j.chb.2007.04.004

[5] Ledbetter A, Larson K (2008) Nonverbal cues in e-mail supportive communication: Associations with sender sex, recipient sex, and support satisfaction. Information, Commun Soc 11:1089–1110

[6] Ledbetter AM, Larson KA (2016) Nonverbal cues in e-mail supportive communication. Information, Commun Soc 11:1089–1110

[7] Moreo a., Romero M, Castro JL, Zurita JM (2012) Lexicon-based Comments-oriented News Sentiment Analyzer system. Expert Syst Appl 39:9166–9180. doi: 10.1016/j.eswa.2012.02.057

[8] Narayanan V, Arora I, Bhatia a (2013) Fast and accurate sentiment classification using an enhanced Naive Bayes model. Int Data Eng Autom Learn Lect Notes Comput Sci 8206:194–201. doi: 10.1007/978-3-642-41278-3_24

[9] Rana TA, Cheah Y-N (2016) Aspect extraction in sentiment analysis: comparative analysis and survey. Artif Intell Rev 46:459–483. doi: 10.1007/s10462-016-9472-z

[10] Rayson P (2002) Matrix: A statistical method and software tool for linguistic analysis through corpus comparison. 194

[11] Riordan M a., Kreuz RJ (2010) Emotion encoding and interpretation in computer-mediated communication: Reasons for use. Comput Human Behav 26:1667–1673. doi: 10.1016/j.chb.2010.06.015

[12] Sánchez-rada JF, Iglesias CA (2016) Onyx : A Linked Data approach to emotion representation. Inf Process Manag 52:99–114

[13] Tang H, Tan S, Cheng X (2009) A survey on sentiment detection of reviews. Expert Syst Appl 36:10760–10773. doi: 10.1016/j.eswa.2009.02.063

[14] Teh PL, Rayson P, Pak I, Piao S, Yeng SM (2016) Reversing the Polarity with Emoticons. In: Métais E, Meziane F, Saraee M, Sugumaran V, Vadera S (eds) Natural Language Processing and Information Systems: 21st International Conference on Applications of Natural Language to Information Systems, NLDB 2016, Salford, UK, June 22-24, 2016, Proceedings. Springer International Publishing, Cham, pp 453–458

[15] Teh PL, Rayson P, Piao S, Pak I (2015) Sentiment Analysis Tools Should Take Account of the Number of Exclamation Marks !!! In: The 17th International Conference on Information Integration and Web-based Applications & Services. ACM New York, NY, USA ©2015, p Artcile 35