

Quantifying Text-based Public's Emotion and Discussion Issues in Online Forum

Siti Z. Z. Abidin, Nasiroh Omar, Muhammad H.M. Radzi and Mohammad B.C. Haron
Faculty of Computer and Mathematical Sciences
Universiti Teknologi MARA
Shah Alam, Selangor, MALAYSIA
sitizaleha533@salam.uitm.edu.my, nasiroh@tmsk.uitm.edu.my,
muhammadhamiz@yahoo.com, bakri891015@gmail.com

ABSTRACT

Currently, an online forum is becoming one of the trends in social media. Social media allow people to communicate interactively so they can share their feelings and opinions. Even though, they are communicating through text, such digital element permits them to express their emotions. The information that is conveyed through texts has the potential to influent people's mind sets. In this paper, we propose to quantify people's perception in an online forum by analyzing their words. The number of occurrences for significant words is counted in order to measure their perception based on emotion and issues that they are discussing. This new approach significantly reduce the time for collecting, processing, and analyzing data, as well as visualizing results, specifically for seeking public's perception and emotion of specific issues.

KEYWORDS

Social media, Emotion, Online forum, Quantifying perception, Issues, Public opinion, Secret Eye.

1 INTRODUCTION

The use of social media has attracted people to communicate and share information freely. People, especially young adults, prefer to choose the online media as their communication tool [1].

The communication allows information to be conveyed through digital elements, especially in the form of text. Even though, they are communicating through text, such digital element permits them to express their emotions [2]. The information has the potential to influent people's mind sets. Any negative influences may lead to social dilemma [3].

In this paper, we propose to quantify people's perception in an online forum by analyzing their words. The number of occurrences for significant words is counted in order to measure their perception based on emotion and issues that they are discussing. There are two cases selected from a Malaysian online forum which is called *forum CARI*. Therefore, the scope of this research focuses on the Malay language. A software tool, *Secret-Eye* [4], is developed to extract all significant words found in the forum for determining the most occurrences words that represent emotion and discussion issues. This new approach significantly reduce the time for collecting, processing, and analyzing data, as well as visualizing results, specifically for seeking public's perception and emotion of specific issues.

The remaining of the paper is organized as follows: Section 2 discusses the related work on gathering public perception, online media and its significant content. Section 3 presents the research methodology, followed by the results and analysis in Section 4. Section 5 concludes the findings.

2 RELATED WORK

Quantifying perceptions are reality and normally measured through an offline survey. Methods in perception quantification have gradually changed from manual paper-based survey to automatic computer-based system. In paper-based survey, the preparation of questionnaire is tedious and time consuming. In addition, the choices of question and answer given to the participants are structured and limited to what is written in the likert-scale survey. Such limitation is not only due to the freedom of expressing opinion or emotion, but also in selecting the type and number of participants. Normally, the survey is sent to the chosen participants and they are expected to understand the questions and return back their responses within a specific time. Unfortunately, the response rate for paper-based survey is lower compared to internet-based survey [5]. In the paper-based survey, the participants' responses are collected for statistical analysis and visualization. These processes are also time consuming and its major drawback is its complexities that require scientific knowledge and skill in statistics which may not be understood by layman audiences [6]. Thus, the computer-based survey provides the opportunities beyond paper-based capabilities in term of information structure, participants, and time to prepare, disseminate, collect and analyze participants' responses.

However, creating and conducting an online survey is a time-consuming task that requires web authoring programs, HTML code, and scripting programs [7]. There may be limitations associated with web survey products and services such as time, space, and number of

respondents. Commonly, such survey is focused on specific participants whom are pre-identified. Nevertheless, the online can offer unlimited respondents covering public opinion which often used to enquire publics' perception on certain products or services.

2.1 Gathering public perception

Perception is defined as the quality of understanding [8]. It includes the theory to understand the surrounding environment [9]. In Information Age, the public can express their perception in digital words. The advancement of Internet technology allows them to express their perception along with emotion in many online platforms such as Weblog, online forum, chatting and emails. Emotion classification can identify the feelings of individuals toward specific events [10]. It is not a trivial task to extract emotional information from the lexical content or meaning of the words in a blog [11]. In addition to words, there are also the emotion icons (emoticons). The emoticons are widely used to represent emotional words [12]. Therefore, public perception can be expressed by using simple but very meaningful icons which have the same meaning with the words that they want to use, such as emotion used in a Weblog computerized system [11].

2.2 Online media

Online media refers to the social media that has been used to do the social interaction with each other online. The example of online media which related to social media is social networking websites, forums and blogs. Social media can be defined as a group of

Internet-based applications that build on the ideological and technological foundations of Web 2.0, which allows the creation and exchange of user-generated content [13].

Currently, the Internet is increasingly popular for survey purposes and it is also used within the higher education environment for handling online information [14]. In education, the self-perception survey can be performed on students' learning to analyze teachers' Personal Style (PS) in the context of science and mathematics teaching [15].

With a new paradigm shift in analyzing information, the self-perception survey method can be transformed into quantification of people's perception. For example, instead of using self-perception survey in quantifying the level of reading comprehension, Omar *et al.* [16] propose a computer-based method. The method can quantify students' quantity of understanding in reading by using a task-based real time data collection in free-response and unstructured style. Thus, it is possible to quantify the students' perception. Similarly, in online media, people are free to express their perception of any issues with the same style. In such case, the digital words that they are using can represent their emotion and discussion issues.

2.3 Emotion and Discussion Issues

It is possible for public to include their perception like in a face-to-face communication. Another way to express emotion is by using emotion icons (emoticons). The emoticons can provide nonverbal replacement, suggestive of facial expression, and may improve the

exchange of emotional information by providing additional social cues beyond what is found in the verbal text of a message [17].

The emoticons are widely used to represent the emotion as compared to using word. The emoticons also are widely used in blogs as user can express their own feelings in their own blog.

As more and more bloggers wish to share their feelings on blogs, some blog service providers start to allow their users to use non-verbal emotional expressions [10]. In recent years, suicide of college students has been a universal phenomenon all over the world. In addition, the number of college students who commit suicide has increased tremendously every year [11]. Therefore, early detection of the emotional conflict can prevent bad consequences.

Moreover, many research works have been conducted to analyze or quantify the emotion through the emoticons. The nearest comparable research work found in the literature is extracting keywords from online Thai texts and classify the keywords according to emotions [20]. Nevertheless, the research focuses on only emotion rather than issues and emoticons. Since this research work is originated from self perception measurement model [19] based on the web text handling model [16], this research is unique and one of its kind.

3 METHODOLOGY

Adapted from the online reading comprehension [16], this research uses similar quantification method to quantify unobtrusively public perception and

emotion found within unstructured online information by using computer-based technology at real time and real life setting.

A software tool called *Secret-Eye* has been developed to read texts from an online forum, *forum CARI*. The forum is the 15th most visited sites in Malaysia, which stands as the second most visited forum site and the first introduced forum in Malay language as the medium of public conversation [18]. This platform is meant for public to discuss any current issues. Anybody can sign as a member of the forum and participate at any time. People can hide their personalities by using nicknames and freely express their opinions.

Secret-Eye is implemented by using C# programming language. It reads and filters the words before clustering them into emotion and facts. At the same time, all the significant words are counted according to the specified cluster. The tool also considers a few emoticons that are clustered as emotion. There are two phases involved in the process; word extraction and quantification.

3.1 Extraction

Inspired by the Self-Perception Measurement Model [19], the extraction of forum's content is performed by reading and filtering all the words. The flow of process during this phase is shown in Figure 1.

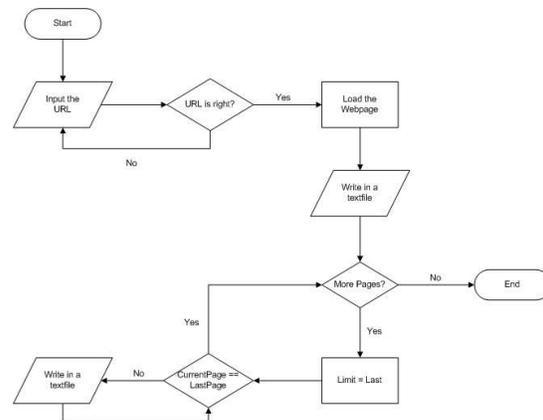


Figure 1. Extraction process

The process starts when the forum's website is found and the number of pages in the forum is determined. All words in the forum are read and stored in a text file for later processing. Therefore, any word that appears in the forum will be processed.

When the extraction process is finished, the filtering process begins as shown in Figure 2.



Figure 2. Filtering process

It is important to select the words that have significant meaning in the discussion and discard any word that acts as tag (in HTML file) or common word. Figure 3(a) illustrates a segment of the forum's content in its original HTML form, while Figure 3(b) shows the same segment after the filtering process.

```
bile tgg krisis kat sana tu kan,.... nmpknya ahkak pink blu ...  
[size=2][color=#999999]dino Post at 7-2-2011 09:08[  
[url=http://mforum1.cari.com.my/redirect.php?goto=findpost&amp;pid=42613825&  
amp;ptid=526936][img]http://mforum1.cari.com.my/images/common/back.gif[img]  
[url][size][quote]
```

(a) Screenshot of the forum's original content

```
bile tgg krisis kat sana tu kan,.... nmpknya ahkak pink blu ...  
dino Post at 7-2-2011 09:08
```

(b) Cleaned content after removing tags

Figure 3. Screenshot of the forum's content

All the significant words are saved as a plain text file. Based on this text file, the unique words and emoticons are manually selected and clustered to be saved in a reference digital database. The database is considered as a case-based or issue oriented corpus to be used in the later phase, the quantification process.

3.2 Quantification

In the quantification process, the words and emoticons in the forum are classified into fact and emotion. The process involves comparing each word in the text with words in the case-based corpus (the reference database). Furthermore, the number of occurrences for each word and emoticon will be stored and ranked. The overall process involves evaluation, classification and calculation. The main purpose is to know how people feel during the discussion and the keywords that can portray the emotion and issues (based on facts) that people are discussing in the forum.

In order to validate the result on the quantification, a quick survey is manually conducted that involves 30 people who are selected at random from

all ages regardless of their background. The majority of these people are of age 20 to 25. This survey will determine whether the suggested keywords are capable to give the clue about the discussion topic.

4 RESULTS AND ANALYSIS

The *Secret-Eye* provides a graphical user interface for the target user to use the system. The examples of target users are investigators or any individual who are interested in the public opinion on any case-based online forum. The main interface of the tool consists of three core sections; URL of the forum, command buttons and display panel. Figure 4 illustrates the main interface of the system.

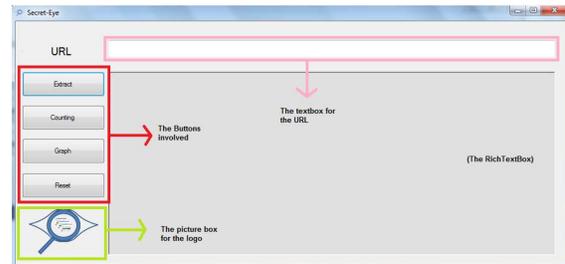


Figure 4. The main interface

At the beginning, all the buttons are disabled except the *Extract* button as the tool will not proceed to the next phase until the extraction process is finished.

Initially, user is expected to enter the URL for the forum before activating the *Extract* button. Once it is started, the tool will extract the archived content of the forum at real time. The extraction starts by reading the web page, step by step from the first page until the last as shown in Figure 5.

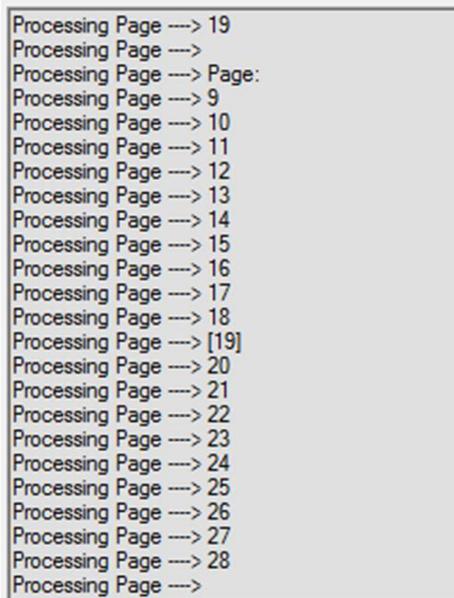


Figure 5. Page extraction

The process continues by performing the filtering phase and all of these are activated by the *Extract* button.

The quantification process begins when the *Count* button is clicked. In this process, all the significant words in the forum are compared to the words stored in the reference database. Figure 6 depicts the word searching and comparison process.

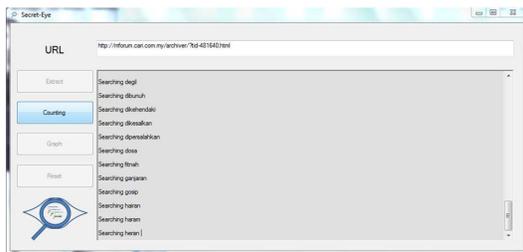


Figure 6. Quantification process

During the process, each significant word will be classified according to the category, either fact or emotion. Concurrently, the frequency of word occurrences is also determined. Figure 7

shows the screenshot of the word occurrence results.

	Word/Emoticon	Count	Category
▶	adil	18	Fact
	akhbar	26	Fact
	al fatihah	5	Emotion
	al-fatihah	0	Emotion
	al-fatiha	0	Emotion
	al-fatihah	5	Emotion
	alfatihah	3	Emotion
	alim	5	Fact
	aman	114	Fact
	amarah	0	Emotion
	anak bongsu	7	Fact
	ancaman	1	Emotion
	baik	132	Fact

Figure 7. Results of quantification

Then, based on the word frequency, the words in each category are ranked and only the top ten words will be visualized by using bar charts. Figure 8(a) displays the graph for emotion while Figure 8(b) shows the top ten words for facts in discussion issues.



(a) Emotion



(b) Discussion issues

Figure 8. Top ten words based on category

Furthermore, the ratio between fact and emotion is calculated in order to identify whether the discussion in the online forum is more towards fact or emotion. The result is displayed in a pie chart as shown in Figure 9.

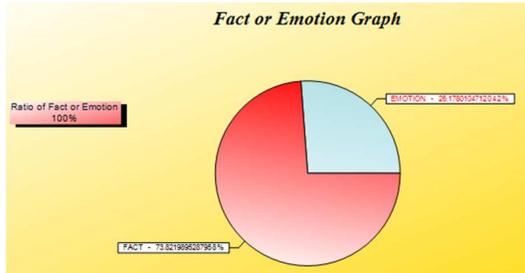


Figure 9. The ratio of fact to emotion

There are two cases used to test the functionality of the tool. The first is *falsely shot innocent local student* and the second case is *tsunami rocks Japan after the massive earth quake*. The results indicate that the tool is able to extract, filter and quantify public perception accordingly. Figure 7 to Figure 9 present the outcome of the first case, while Figure 10 shows one of the results of the second case.

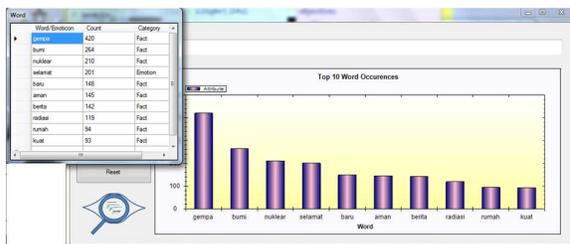


Figure 10. Top ten word occurrences

At this point, the tool has fulfilled the functionality requirements. Furthermore, it has undergone the usability testing. Based on a paper-based survey with 32 participants, 28 of them are able to identify the discussion topic correctly when they refer to the graph of top ten significant words. The outcome of the

survey indicates that the proposed method of quantifying virtual public's perception is feasible to capture the clue in the online forum discussion. Figure 11 illustrates the outcome of the usability test based on the percentage of guesses.

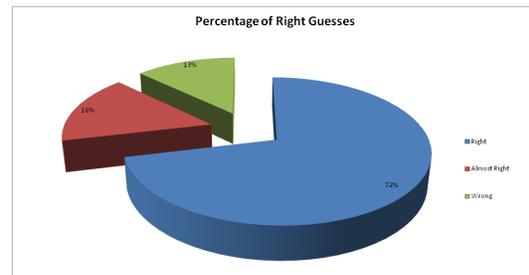


Figure 11. Results of usability test

Based on the results, 72% (23 respondents) are able to identify correctly that the top ten word occurrences relate to main subject in the first case. The other 16% (5 respondents) can give the right guess but with wrong spelling. They have the clue about the case but they miss the details. The rest of 13% (4 respondents) have no idea about the discussion issue.

5 CONCLUSION

In this paper, we have presented the first potential attempt to quantify public perception and their emotion of specific issues through capturing top ten significant words that are categorized into facts and emotion. The scope of this research focuses on the Malay text extracted from a case based online forum. While most methods in quantifying public's perception are performed through survey-based methodology, we have shifted the paradigm by using a real-time data

collection on an online-based case. By using the online forum, data collection on the public's opinion can be performed unobtrusively. At the same time, we argue that most people, who involve in the online discussion, will express their ideas freely and naturally.

We conclude that it is feasible to capture the popular discussion issues that the public is chatting and arguing in the online forum. This new approach significantly reduce the time for collecting, processing, and analyzing data, as well as visualizing results, specifically for seeking public's perception and emotion of specific issues. It has the potential to be strengthened in helping any individual or organization who seeks public's opinion in making decision. In the future, the content of the database dictionary will be expanded to include many languages and issues.

6 ACKNOWLEDGEMENT

We would like to thank the Universiti Teknologi MARA Malaysia for the financial support (Excellence Fund).

7 REFERENCES

1. Rahman, H.F.A., Omar, N., Abidin, S.Z.Z., Mahmud, Z., Rosli, M.M.: Visualizing Patterns of Online Media Preference Based on Young Adults Lifestyle, Proceedings of the 15th WSEAS International Conference on Computers, (2011)
2. Idrus, Z., Abidin, S.Z.Z., Hashim, R., Omar, N.: Social Awareness: The Power of Digital Elements in Collaborative Environment, WSEAS Transactions on Computers, Vol. 9(6), pp. 644—653, (2010)
3. Ottaway, M., Hamzawy, A: Protest Movements and Political Change in the Arab World. 2011 [cited; Available from: http://www.carnegieendowment.org/files/OttawayHamzawy_Outlook_Jan11_ProtestMovements.pdf]
4. Abidin, S.Z.Z., Omar, N., Radzi, M.H.M., Haron, M.B.C: Secret-Eye: A Tool to Quantify User's Emotion and Discussion Issues through a Web-Based Forum, In: Zain, J.M. et al. (eds.) ICSECS 2011. Part III, CCIS 181, pp. 455--462. Springer, Heidelberg (2011).
5. Cobanoglu, C., Warde, B. And Moreo, P.J.: A Comparison of Mail, Fax and Web-based Survey Methods. In: International Journal of Market Research (2001)
6. Harrison, C.: Postmodern Research and e-learning: Anatomy and Representation. In : European Educational Research Journal. pp.80-93. (2006)
7. Wright, K. B.: Researching Internet-based Populations :Advantages and Disadvantages of Online Survey Research, Online Questionnaire Authoring Software Packages, and Web Survey Services. Journal of Computer-Mediated Communication, 10(3), article 11 (2005)
8. Hornby, A. S. (Ed.): Perception. In: Oxford Fajar dictionary, pp. 1335. Shah Alam (2000)
9. Sun Libo, L. Y.: The Hierarchical Perception Model for Crowd Simulation. In: Sixth International Conference on Computer Graphics, Imaging and Visualization. pp. 106-111 (2009)
10. Yang, C., Lin, K. H.-Y., & Chen, H.-H.: Emotion Classification Using Web Blog Corpora. In: IEEE/WIC/ACM International Conference on Web Intelligence (2007)
11. Li, J., & REN, F.: Emotion Recognition from Blog Articles. In : International Conference on Natural Language Processing and Knowledge Engineering. pp. 1-8. NLP-KE '08, Beijing (2008)
12. Thompson PA, F. D.: Effects of Pictographs and Quoting on Flaming in Electronic Mail. In : Computers in Human Behavior, pp. 225-243 (1996)
13. Kaplan, A., & Haenlein, M.: Users of the World, Unite! The Challenges and Opportunities of Social Media. Business Horizons, 53(1), 59—68. (2010)
14. Leu, D. J., Neag, J. and Neag, M.: Expanding the Reading Literacy Framework of PISA 2009 to Include Online Reading Comprehension. In: A working paper commissioned by the PISA 2009 Reading Expert Group (2009)

15. Ben-Chaim, D., & Zoller, U. : Self-Perception versus Students' Perception of Teachers' Personal Style in College Science and Mathematics Courses. In: Research in Science Education. pp. 437-454. (2001)
16. Omar,N., Higgins, C., Harrison, C. and Campo Millan,D.: Evaluating Real-time Online Research Data (RORD) and Verbatim Quotient Detection (VQD): Low Inference Tools to Monitor Outcomes of Unconstrained Authentic Internet Research. In: Evaluating Real-time Online Research Data (RORD) and Verbatim Quotient Detection (VQD), pp. 502-506. Netherland (2006)
17. Derks, D., Bos, A. E., & Grumbkow, J. V.: Emoticons in Computer-Mediated Communication: Social Motives and Social Context. *CyberPsychology & Behavior* (2008)
18. Top Sites in Malaysia, <http://www.alexa.com/topsites/countries/MY>
19. Omar, N., & Abidin, S. Z.: Towards Measuring Self-Perception in Disseminating Information. In: Proceedings 2010 International Conference on Information Retrieval and Knowledge Management, pp 147-151. CAMP'10,Shah Alam, Malaysia (2010)
20. Inrak, P. & Sinthupinyo, S.: Applying Latent Semantic Analysis to Classify Emotions in Thai Text. In: International Conference on Computer Engineering and Technology, pp.450-454.(2010)