

Measuring Malaysian Public Perception with Secret-Eye

Siti Zaleha Zainal Abidin,^{1,2} Nasiroh Omar² & M. Hamiz M. Radzi²

¹Faculty of Media and Communication Studies,

²Faculty of Computer and Mathematical Sciences,

Universiti Teknologi MARA

ABSTRACT

People's perception and emotion are usually measured through conducting a survey research. With the rigid multiple choice or lickert scale selection of answers, participants have to respond within a given time frame to give their feedback on a certain specified issue. In this paper, we present our research through collecting data regarding people's perception and emotion through an online forum. We quantify the data and perform data classification based on issues and emotion. Data collection are carried out by using an implemented software tool called Secret-Eye which is capable of extracting all significant words in the forum unobtrusively at real time real life basis. The classified data are ranked according to their frequency of existence. The discussion on the tsunami disaster in Japan that happened in March 2011, is selected for this study. It is beneficial to have the results on the perception analysis as public's feelings and facts on the discussion are expressed naturally. Thus, the results indicate that it is feasible to extract people's true means of perception through social media.

Keywords: Data collection, Information gathering, Online forum, Public perception, Secret-Eye.

Introduction

Perception is defined as the quality of understanding (Hornby, 2000). People's perception is usually based on their own attitudes and feelings in various situations (Bem, 1967). The attitudes and feelings can be expressed through spoken, written or gestures. In others' view, their

attitudes and feelings represent perception and emotion. These are usually measured through conducting a survey research. With the rigid multiple choice or lickert scale selection of answers, people have to respond within a given time frame to give their feedback on a specific issue. The expression of emotion in face-to-face communication depends mostly on nonverbal reminders. These cues are not present in text-based computer-mediated communication (CMC). Emoticons provide nonverbal replacement, suggestive of facial expression, and thus, it may improve the exchange of emotional information by providing additional social cues beyond what is found in the verbal text of a message (Derks et al., 2008).

In this paper, we present our research through collecting data regarding people's perception and emotion through an online forum. We quantify the data and perform data classification based on issues and emotion. Data collection are carried out by using an implemented software tool called Secret-Eye (Abidin et al., 2011) which is capable of extracting all significant words in the forum unobtrusively at real time real life basis. The classified data are ranked according to their frequency of existence. The discussion on the tsunami disaster in Japan that happened a few months ago is selected for this study. It is beneficial to have the results on the perception analysis as public's feelings and facts on the discussion are expressed naturally. Thus, the results indicated that it is feasible to extract people's true means of perception through social media.

This paper is organized as follows: Section 2 discusses about social media and how data on the social media can be gathered unobtrusively for content analysis, while Section 3 presents the methodology of this research. Section 4 gives the results and analysis before the concluding remarks in Section 5.

Social Media

Nowadays, social media has become one of the popular tools for people to stay connected to share opinion, information, work, stories and many more. Among all, the Internet has attracted people since it helps the interaction becomes easier and provides the new alternative to communicate with other people. The term online media is often used to refer to the Internet based media. The few examples of online media which relate to social media are social networking websites, forums

and blogs. Thus, social media can be defined as a group of Internet-based applications that are built on the ideological and technological foundations of Web 2.0, which allows the creation and exchange of user-generated content (Andreas & Haenlein, 2010).

Yahoo! Messenger, Google Talk and Live Messenger are the examples of Computer-Mediated Communication (CMC). CMC is defined as any communicative transaction that occurs through the use of two or more networked computers such as instant messaging and chat room (Ahern et al., 1992). The frequent CMC users are mostly youngsters. In order to add enjoyment and humor, emotional icons (emoticons) are used on Bulletin Board System (BBS) (Derks et al., 2008). Emoticons can express the emotion of users and it is possible to have different emoticons to convey the same meaning (Derks et al., 2007), even though it is usually measured through conducting a survey. In addition to emoticons, information in online media can be in many forms including texts, images and graphics.

Receiving information, synthesizing the information and communicating the information to others are gradually more important currently (Leu et al., 2009). Normally, the information is neutral and it is possible to use the information to manipulate people's perception (Lawrence, 2009). Therefore, it is possible to use the internet as the platform to disseminate wrong information which can lead to threat (Omar and Abidin, 2010).

Web Content Analysis

A web forum is one of the online media that provide a platform for any Internet users around the world to communicate with each other and express their opinions (Yang and Ng, 2008). As web forum has become an enormous collection of highly valuable opinions and commentaries, more and more researchers are interested in analyzing the content of a forum. However, most of them pay attention to the forum reviews rather than the posts themselves (Shi et al., 2009). In addition, more than two-thirds of the global online population visit and participate in social networks and blogs. In fact, social networking and blogging account for nearly 10% of all time spent on the Internet (Benevenuto et al., 2009).

Normally, the forum users interact with each other in a thread in a forum when they have a common interest. A web forum is a virtual platform for expressing personal and communal opinions, comments, experiences, thoughts, and sentiments (Yang and Tobun, 2007).

Moreover, Shi et al. (2009) also agreed that the web forum is becoming an increasingly popular online form of conversation and communication. The assessment of performing content clustering on forum's interactive discussions has two folds. The first is to identify and group similar threads together and hence to abstract the topics or themes from all clusters and the second one is to unveil the ideological similarity between forum participants who may or may not have direct interaction.

The asynchronous nature of web forum allows multiple participants to discuss one topic in parallel. Some more, nowadays, it is no mystery that terrorists and cyber-criminals are looking for uncovered means to seek peers with the same interests such as internet forums, blogs, and social networks, where they can share their feelings and interests with other that support their action (L'Huillier and Alvarez, 2010). Any participant in forum discussion is able to post a thread and start the discussion on its topic (Yang and Tobun, 2007). Because of this self-interest-oriented posting mechanism, it is possible that the topic of a thread may be unique among all other threads in forums.

Measuring Emotion

Emotion classification allows the feelings of individuals to be identified towards specific events (Yang, 2007). It is not a trivial task to extract emotional information from the lexical content or meaning of the words in a blog (Li et al., 2008). Therefore, normally people will express their emotion in words. Nevertheless, since the growth of internet, there are a lot of platforms for them to express the emotion. The blogs and instant messaging have attracted users' interest. The Emotion Icons (emoticons) are widely used to represent the emotion rather than using word to symbolize it (Thompson, 1996). The emoticons also are widely used in blogs as users can express their own feelings in their own blog.

Methodology

Based on the Model of Web Text Reading Comprehension (Omar et al., 2004), this research aims to investigate Malaysians public perception on the use of the words (in Malay language) in an online forum. The case study is *Forum CARI* which is ranked as among the top 15 popular web sites in Malaysia (Top Site, 2011). Within the forum, comments from the public are retrieved based on the issue of tsunami disaster in Japan.

A computer program is designed (namely *Secret-Eye*) and developed in a high level programming language to read all words in a selected topic in the forum which usually consists of thousands of words. The program that sits on a personal computer will be linked to the specified server that contains the specified forum. A database is also needed to keep all the significant and meaningful words used in this research (Figure 1). Since the program will only investigate words, all emoticons are converted into text when the program performs parsing of the forum's content.

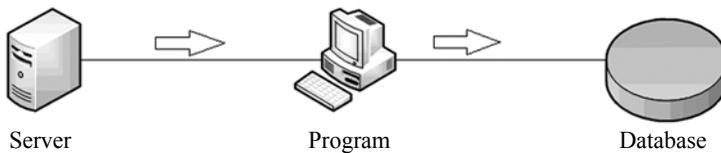


Figure 1: Software Design Overview

In order to perform the measurement of perception, the design for the software tool involves several steps as follows:

- a. *Database preparation* – A collection of significant words are classified either *emotion* or *fact* and they are stored in a database. The word determination is performed manually by browsing through the forum content. More than 1000 significant keywords are identified and selected as *corpus*.
- b. *Web data extraction* – When the address location of the selected forum is identified, words are extracted from the web that has more than hundred pages according to interest of the public to discuss about the issue.
- c. *Word Filtering* – Since the web pages contain many tags along with the significant words, the software tool will filter the forum content so that only the meaningful words are going to be processed.
- d. *Word counting based on frequency* – While doing the filtering, this software tool makes the comparison with the database to check the classification of the word while counting its frequency of occurrence. The results are stored in a table that keeps the information about the significant keyword and its related count and category (classification).
- e. *Displaying emotion to fact ratio* – Based on the created table, a pie chart is produced to give the ratio of fact and emotion being discussed

in the forum. This gives the measurement either the public involved in the discussion are leaned towards emotion or facts.

- f. *Displaying top ten highest word frequency based on all words occurrence, fact, emotion* – A graphical presentation of the word ranking is provided to give the overall feelings and facts about the issue discussed.

With the above steps, the public’s perception can be measured to give the concluding remarks on how the public perceive the issue.

Results and Analysis

On 11th March 2011, Japan was struck by an 8.9 Richter magnitude scale earthquake off its north-east coast, triggering a huge tsunami that washed away cars and buildings along the coast. This tragedy is discussed widely by the forum users as this is one of the greatest earthquake ever happened in Japan for 100 years. The words in the forum are analyzed to form the top ten word occurrences. The analysis of the is plotted in the Figure 2.

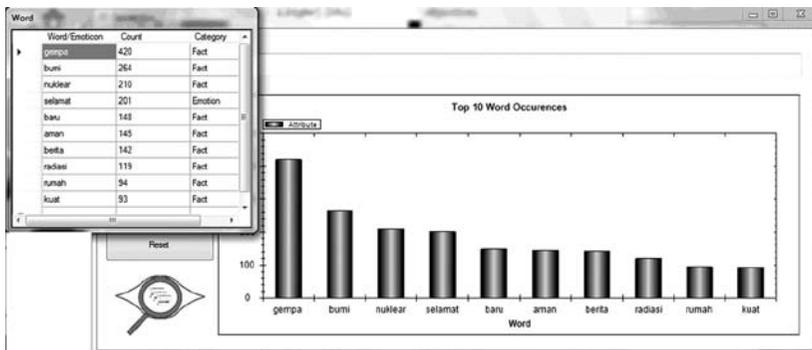


Figure 2: Top Ten Word Occurrences

The words *gempa*, *bumi*, and *nuclear* are the most frequently occurred. The composition of words *gempa* and *bumi* brings the meaning of *earthquake*. The result shows that the people are concerned about the earthquake and the nuclear. The rest of the frequent significant keywords reflect the elements in the scenario. In addition, the results of top ten word occurrence and emoticons to reflect people’s emotions is also produced. Figure 3 illustrates the analysis which presents the feeling of distress

is more than happiness. However, the single keyword used seems not enough to represent emotion because the emotion can be interpreted in various ways and must be viewed within its context.

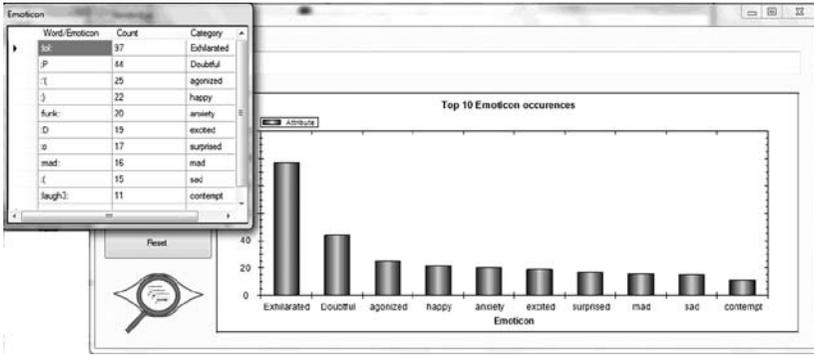


Figure 3: Top Ten Word Occurrence for Emotion

Figure 4 shows the ratio between words of fact and emotion. Based on the results, 74.78% of the words represent fact and the rest, 25.22% signify public’s emotions.

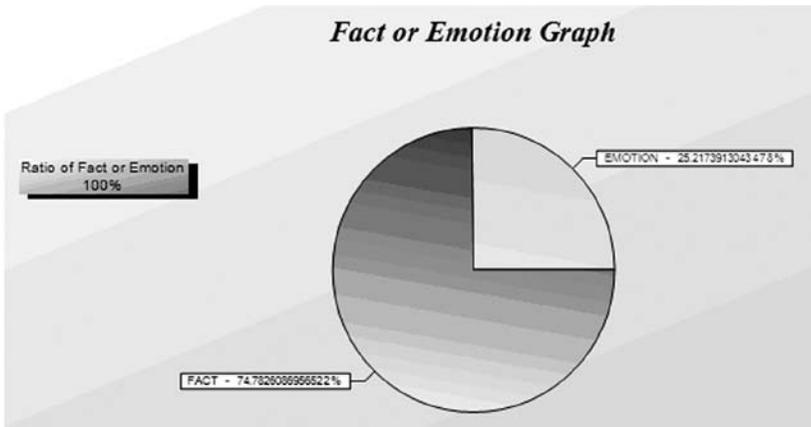


Figure 4: A Chart on Fact vs. Emotion

Therefore, the results demonstrate that most people discuss about fact rather than expressing their feelings when discussing about the tsunami issue.

Conclusion

In this paper, we have presented a computer-based approach in measuring public's perception that are conveyed naturally and freely through the online social media. With the significant keywords extracted automatically and unobstrusively at real time, it is feasible to determine issues and emotion in the discussion. Even though, this approach involves only text elements and a few emoticons, it has the potential to provide a flexible mechanism in collecting data that are related to people perception instantly.

References

- Abidin, S.Z.Z., Omar, N., Radzi, M.H.M. & Haron, M.B.C. (2011). Secret-Eye: A Tool to Quantify User's Emotion and Discussion Issues through a Web-based Forum, ICSECS 2011, Part III, CCIS 181, pp. 455-462. Springer-Verlag.
- Ahern
- Andreas
- Bem, D. J. (1967). Self-Perception: An Alternative Interpretation of Cognitive Dissonance Phenomena. *Psychological Review*, 183-200.
- Benevenuto
- Derks, D., Bos, A. E. & Grumbkow, J. V. (2008). Emoticons in Computer-Mediated Communication: Social Motives and Social Context. *CyberPsychology & Behavior*.
- Hornby, A. S. (Ed.). (2000). Perception. In: *Oxford Fajar dictionary*, pp. 1335. Shah Alam.
- Lawrence, D. (2009). Measures of Effects. Information Operation Course. FKPM, UiTM, Malaysia.

Leu, D. J., Neag, J. & Neag, M. (2009). 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.

Li. (2008).

L'Huillier & Alvarez. (2010).

Omar, N. & Abidin, S. Z. (2010). 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.

Omar. (2004).

Shi. (2009).

Thompson. (1996).

Top Sites in Malaysia, <http://www.alexa.com/topsites/countries/MY>

Yang. (2007).

Yang & Ng

Yang & Tobun. (2007).