Secret-Eye: A Tool to Quantify User's Emotion and Discussion Issues through a Web-Based 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

Abstract. Receiving, synthesizing and communicating information to others are gradually more important. Information can be in many forms, and text can represent a lot of information that includes people's emotion. Emotion is usually measured through conducting a survey. For digital information, emotion can be expressed in words or emoticons. In this paper, we propose to quantify the user's perception in a web-based forum through analyzing words used by users. By determining the number of occurrences for each significant word in the forum, we measure the user's perception based on emotion and issues that they are discussing. Therefore, we implement a software tool called Secret-Eye that provides a platform to extract words in any web-based forum. The tool is implemented using C# 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. This approach is significant in identifying public opinion that can collect people's perception and discussion issues unobtrusively through case-based online forum. At this stage, the scope of this research is focused on a Malay language which will be extended to be used in other languages. The tool is important to Malaysian organizations, individuals and investigators who seek for public opinion as well as emotion through online forum. There are many tools for collecting public's perception, however, what is novel in this research is the use of case-based real time data collection method that highlights potential areas for using computer-based technology in quantifying public emotion and perception.

Keywords: Quantifying perception, Web-based, Real time data collection, Word, Emoticon, Emotion, Issues, Public opinion, Secret Eye.

1 Introduction

The use of internet allows ordinary people from all ages to express their opinions and feelings on any issue at their own convenience. People at young age are regularly expressed their views in complex sentences via e-questionnaires, forums or Weblogs. Moreover, they often choose to use natural expressions, specifically, in the form of "youth words" (Internet slang) and textual emoticons [1]. Although the people's

J.M. Zain et al. (Eds.): ICSECS 2011, Part III, CCIS 181, pp. 455–462, 2011.

communication and comments are considered normal conversation and parts of modern social network, their perception may leads to threats [2], due to the fact that the perception of a person could be manipulated [3]. Originated from self-perception approach, it is possible to quantify people's opinion based on the texts that they use in their conversation. Self-perception is an awareness of the characteristic that constitute one's self. It means, people will decide on their own attitudes and feelings from watching themselves behave in various situations [4]. In addition, self perception is about looking a glass-self, how people think they may appear to others; how they think other people may evaluate their appearance and the resulting shame or pride others' feel [5]. In simpler words, self-perception can be defined as how people conduct themselves in different states and how other people evaluate about it.

Previously, the investigation on self-perception is carried out by using paper-based survey and people who participate are aware of the evaluation process and purpose of the survey. Since then, many tools for collecting public's self-perception have been advanced in parallel to the advancement of technology. Instead of conducting paper-based self-perception surveys, researchers can conduct computer-based self-perception survey. A review of literature shows that currently there is no existing system to quantify the perception and emotion of the received information via online media and produce instant feedback in a real-life setting.

This paper proposes a method to quantify people's perception in a web-based forum through analyzing words used by users. By determining the number of occurrences for each significant word in the forum, people's perception is quantified based on emotion and issues that they are discussing. Therefore, a software tool is implemented called *Secret-Eye* that provides a platform to extract words in any web-based forum. The tool is implemented using C# 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. This approach is significance in identifying public opinion that can collect people's perception and discussion issues unobtrusively through case-based online forum. At this stage, the scope of this research is focused on Malay language which will be extended to be used in other languages. The tool is important to Malaysian organizations, individuals and investigators who seek for public opinion as well as emotion in online forum. The novelty of this research is the use of real time data collection method that highlights potential areas for using computer-based technology in quantifying case-based public emotion and perception.

This paper is organized as follows. Section 2 discusses on research of quantifying people's perception. Next, Section 3 presents the framework and methodology of the software tool (*Secret-Eye*). Section 4 discusses the result and analysis of this research before the concluding remarks in Section 5.

2 Quantifying Perception

Perception is defined as the quality of understanding [6]. It includes the theory to understand the surrounding environment [7]. In Information Age, people usually

express their perception in digital words. The advancement of Internet technology allows people 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 [8]. It is not a trivial task to extract emotional information from the lexical content or meaning of the words in a blog [9]. In addition to words, there are also the emotion icons (emoticons). The emoticons are widely used to represent emotional words [10]. Therefore, people's 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 quantifying emotion in a Weblog using a computerized system [9].

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 [11]. 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 require scientific knowledge and skill in statistics which may not be understood by layman audiences [12]. 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.

Currently, the Internet is increasingly popular for survey purposes and it is also used within the higher education environment for handling online information [13]. 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 [14]. 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.* [15] 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. Similarly, in online media, people are free to express their perception of any issues with the same style . However, it is possible to include their perception embedded with the cues to express their emotion like in a face-to-face communication, which depends mostly on nonverbal reminders. Another way to express emotion is by using 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 [16]. The emoticons can also be represented in words. Adapted from the online reading comprehension [15], this research uses similar quantification method to quantify unobtrusively people's perception and emotion through unstructured online information by using computer-based technology at real time and real life setting.

3 Secret-Eye Work Flow

In this research, a software tool for quantifying public's perception called *Secret-Eye* is implemented. As a case study, a Malaysian online forum, namely *forum CARI*, a social media platform is used. This platform is meant for people 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. The reason of choosing this particular website is due to its popularity as the 15th most visited sites in Malaysia, the second most visited forum site and the first introduced forum in Malay language as the medium of conversation [17]. There are two phases in the implementation of the tool. The first phase is word extraction and the second phase is quantification process.

3.1 Extraction

Inspired by the Self-Perception Measurement Model [2], the extraction of forum's content is performed by pulling out all the words. The flow of process during this phase is shown in Fig 1. 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.

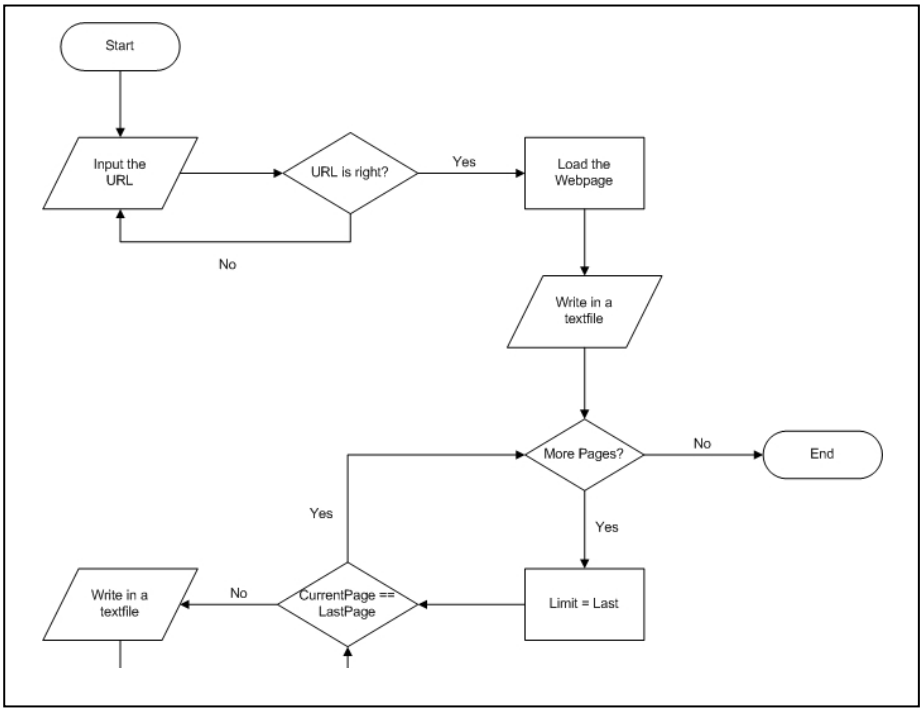


Fig. 1. Extraction Process

When the extraction process is finished, 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. 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 saved in a digital database, as part of case-based corpus for the language used in the forum.

3.2 Quantification

In the quantification process, the words and emoticons in the database are classified into *fact* and *emotion*. The process involves comparing each word in the text with words in the case-based corpus. 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 design and implementation of *Secret-Eye* is in *C#* programming language. The tool 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 results of this research are visualized in five graphical forms including bar charts and pie chart. Fig.2 displays the screenshot of *Secret-Eye* that presents the content of a text file with all the significant words after the tags and common words cleaning process, while Fig.3 shows the process of word search and classification in relation to the corpus in the database.

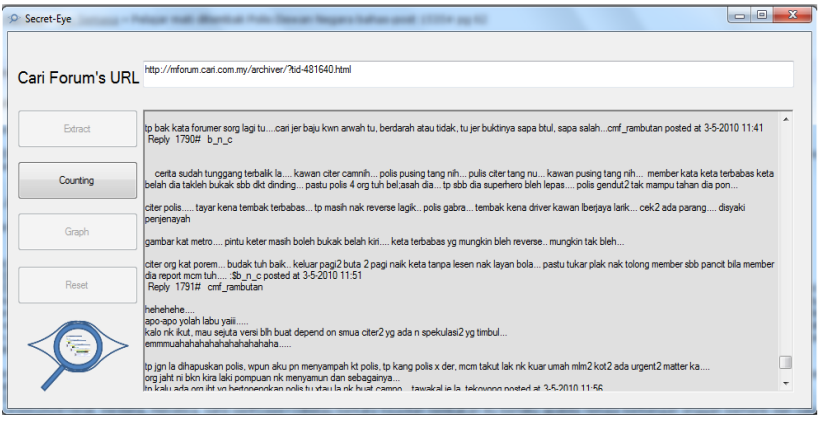


Fig. 2. Display of the filtered file

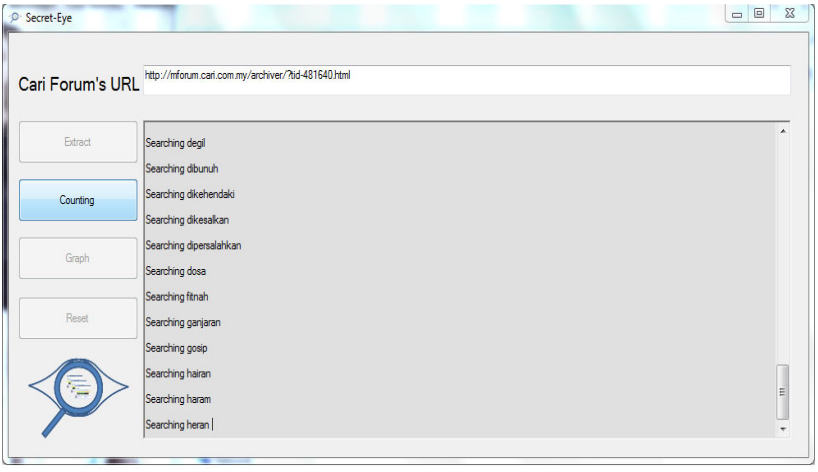


Fig. 3. The process of searching word occurrences in corpus

Fig.4 illustrates the output that contains the table of word occurrences after comparing the selected word with the case-based corpus dependent database. The result is presented in a bar chart of the top ten keywords for determining discussion issues.

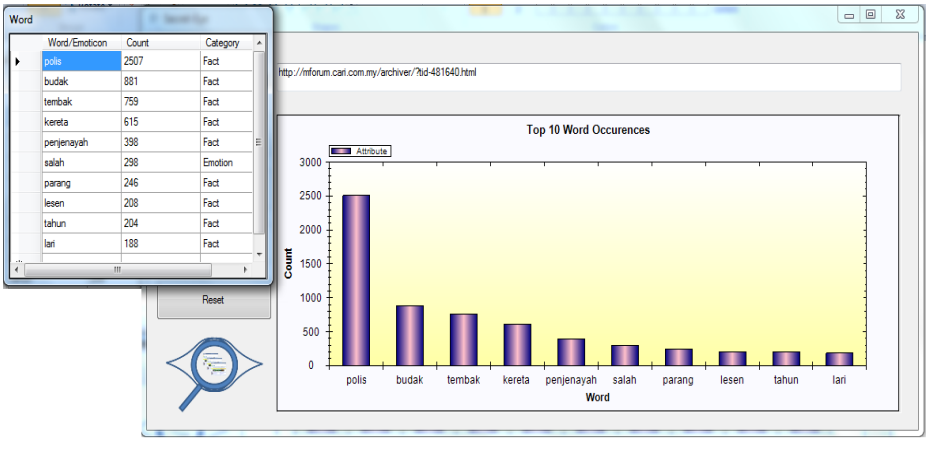
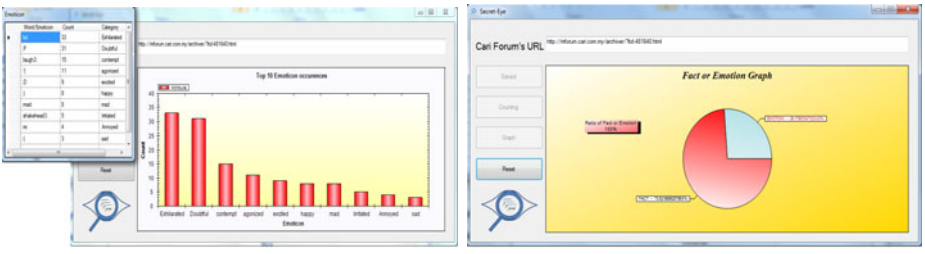


Fig. 4. The occurrences of top ten keywords on discussion issues

Fig.5 shows the two results of the quantification in a pie chart form: (a) top ten keywords for presenting people's emotion; and (b) the ratio of words for emotion in relative to facts. This ratio is calculated to check whether the discussion in the online forum is more towards emotion or facts. Therefore, this value can indicate whether most people are emotional or serious in the discussion.



(a) Top 10 keywords for emotion (b) The ratio of Fact to Emotion

Fig. 5. Results of top ten keywords for emotion and the ratio of fact to emotion in the forum

The other two results present the bar chart of top ten keywords for emotion based on emoticons and top ten keywords on facts. Based on the quick 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 keywords. 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.

Since this work is originated from self perception measurement model [2] based on the web text handling model [15], this research is unique and one of its kind. Thus, the nearest comparable research work found in the literature is extracting keywords from online Thai texts and classify the keywords according to emotions [18]. Nevertheless, the research focuses on only emotion rather than issues and emoticons.

5 Conclusion

In this paper, we have presented the first potential attempt to quantify people's perception and their emotion of specific issues through capturing top ten keywords that are categorized into emotion and facts through case based online forum, performed at real time and real life situation. While most methods in quantifying people's perception are performed through survey-based methodology, we have shifted the paradigm by using a real-time data collection on a web-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.

The tool, named as *Secret-Eye* can extract all the words in the specified online forum and perform evaluation and analysis before producing four categories of top ten keywords; emotion, emoticon, facts, and occurrences. We conclude that it is feasible to capture the popular discussion topic 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.

Acknowledgements

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

References

1. Itoh, M.: Contextual analysis of Complex Sentences Expressing Sentiments. In: 2009 Eight International Symposium on Natural Language Processing, pp. 5–10. IEEE, Los Alamitos (2009)
2. Omar, N., Abidin, S.Z.: Towards Measuring Self-Perception in Disseminating Information. In: Proceedings 2010 International Conference on Information Retrieval and Knowledge Management, CAMP 2010, Shah Alam, Malaysia, pp. 147–151 (2010)
3. Lawrence, D.: Measures of Effects. Information Operation Course. FKPM, UiTM, Malaysia (2009)
4. Bem, D.J.: Self-Perception: An Alternative Interpretation of Cognitive Dissonance Phenomena. Psychological Review, 183–200 (1967)
5. Self Perception, http://webspace.ship.edu/ambart/PSY_220/Self_percol.htm
6. Hornby, A.S. (ed.): Perception. Oxford Fajar Dictionary, p. 1335. Shah Alam (2000)
7. 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)
8. 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)
9. Li, J., Ren, F.: F. Ren Ren Emotion Recognition from Blog Articles. In: International Conference on Natural Language Processing and Knowledge Engineering, IEEE NLP-KE 2008. Beijing, pp. 1–8 (2008)
10. Thompson, P.A., Foulger, D.A.: Effects of Pictographs and Quoting on Flaming in Electronic Mail. In: Computers in Human Behavior, pp. 225–243 (1996)
11. Cobanoglu, C., Warde, B., Moreo, P.J.: A Comparison of Mail, Fax and Web-based Survey Methods. International Journal of Market Research (2001)
12. Harrison, C.: Postmodern Research and e-learning: Anatomy and Representation. European Educational Research Journal, 80–93 (2006)
13. Leu, D.J., Neag, J., 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)
14. Ben-Chaim, D., Zoller, U.: Self-Perception versus Students' Perception of Teachers' Personal Style in College Science and Mathematics Courses. Research in Science Education, 437–454 (2001)
15. Omar, N., Higgins, C., Harrison, C., 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), Netherland, pp. 502–506 (2006)
16. Derks, D., Bos, A.E., Grumbkow, J.V.: Emoticons in Computer-Mediated Communication: Social Motives and Social Context. CyberPsychology & Behavior (2008)
17. Top Sites in Malaysia, http://www.alexa.com/topsites/countries/MY
18. 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)