

EMOTION DETECTION AND OPINION MINING FROM STUDENT COMMENTS FOR TEACHING INNOVATION ASSESSMENT

Angelina Tzacheva¹ and Akshaya Easwaran²

¹Department of Computer Science, University of North Carolina at Charlotte,
Charlotte, North Carolina – 28223, USA

²Department of Information Technology, University of North Carolina
at Charlotte, Charlotte, North Carolina - 28223, USA

ABSTRACT

Students can provide their opinions, comments, or suggestions about a course, course instructor, study environment, and available resources using the course evaluation at the end of every semester. This helps the course professors and other college authorities make appropriate changes or continue a particular approach to get the best experience in classrooms. These course evaluations are in both quantitative and qualitative forms. In quantitative feedback the evaluation is performed in terms of measurable outcomes and include a Likert-type scale to capture the level of agreement and disagreement. In qualitative feedback the students can convey their feelings, opinions or suggestions about the course, the course instructor, or their overall thoughts/comments towards the course. The qualitative feedbacks provide freedom for the students to express their honest thoughts on a course. The data collected in the qualitative form provides deeper insight into a student's emotional state. In this work we focus on mining the qualitative student feedbacks and analyzing the student sentiments. We also analyze the efficiency of Light Weight teams and Flipped Classroom approach which are Active Learning methods. Results show that the implementation of these Active Learning methods is linked with increased positivity in student emotions.

KEYWORDS

Sentimental Analysis; Qualitative Feedback; Active Learning Approach; Emotion Mining; Flipped Classroom; Light Weight Teams; Emotion Detection; Opinion Mining.

1. INTRODUCTION

Student evaluation has become increasingly recognized as a valid empirical foundation for the institutional assessment of all matters related to teaching: teachers, teaching and academic programs as described by Darwin [1]. These evaluation results apart from providing meaningful insights to improve student learning experience also help the college administration in making faculty related decisions like future employment, retention, and promotion. Almost all educational institutions these days collect the course evaluation feedbacks from students at the end of the semester. These evaluations help the instructor as well as the college administration to know what the students felt about the course, the course instructor, the assessment techniques implemented in the class and the course environment in general. The evaluation usually consists of two types of questions: Quantitative and Qualitative type questions. The Quantitative format questions are Likert-type questions in which the students can respond in the scale of 1 to 5, starting with Strongly Agree - 1, Agree - 2, Neutral - 3, Disagree - 4, and Strongly Disagree - 5. The Qualitative format questions are open ended questions where the students have the freedom to write their opinions, suggestions, or feelings in a free style manner. The student evaluation at

recent times is generally linked to two closely related activities [1]. The first was the introduction of a student ratings form the second was the release of a study on the design of student ratings ([2], [3], [4]).

Data mining is the process of using special software to look at large amounts of computer data in order to find out useful information [5]. This information in the form of patterns are then analyzed to extract details that can help improve the existing experience in an area. Authors Spooren et al. [6] claim that there are three main purposes for using student evaluations. They are a) to improve teaching methodology and/or quality, b) to serve as input for tenure/promotion decisions, and c) to demonstrate the evidence of institutional accountability in terms of resources and environment provided. Mining the student evaluations data is gaining popularity and importance in the recent days to improve students' educational outcomes and to ensure quality preparation of all students to life and work. The data generated by student evaluation is also becoming metric for ranking various universities and also a popular guide for students that rate comparative institutional quality [1]. In higher education environments, student evaluation is being increasingly employed to address the imperatives around quality improvement, quality assurance and performance management [1].

Emotion mining techniques are used by many leading technology companies like Google, Microsoft, IBM, SAP, etc. [7] to build their own in-house industry activities like product improvement, and competitor analysis. Over the last few years, almost any product or service used by people are reviewed by them for its quality/effectiveness. Primary examples are food reviews, e-commerce product reviews and trip reviews. Such reviews help the businesses to understand where their product/service stands among their customers and also help them in making business decisions related to it.

According to Merriam-Webster Dictionary, Emotion is defined as “a conscious mental reaction (such as anger or fear) subjectively experienced as strong feeling usually directed toward a specific object and typically accompanied by physiological and behavioral changes in the body” [8].

The authors Ortony and Turner discuss different basic emotion models proposed by various theorists since 1960 [9]. In this work we use the National Research Council - NRC Emotion lexicon for emotion labeling of the text [10], [11].

Recent growth and development of big social media platforms like Twitter, Facebook, Instagram etc. and personal blog sites generate huge amount of unstructured data. Companies providing products and services collect customer satisfaction data through surveys. Educational institutions collect data from students on their opinion for their learning experience in the courses. Analysis of this data may provide insights into the opinion of people as well as their feelings towards certain subjects, products or services. The process of extracting valuable data from opinions of people to assess their feelings and thoughts is known as opinion mining. Mining the opinions of people has applications in several areas: understanding what people like or dislike is critical for making informed business, educational and political decisions.

Emotion is a primary concern in younger generation students that have major impact on the productivity in school. In this paper we mine the qualitative student feedback results from end-of-semester course evaluations to identify the student emotions. We understand whether incorporation of Light Weight teams [12], [13], and Flipped Classroom techniques [14] helped students during the course for the time period 2013 to 2020 from the identified student emotions.

Reminder of this paper is organized as follows: section II talks about the related work in the area, section III the methodology in data extraction and emotion labeling followed by Experiments and results in section IV and Discussion and Conclusion in section V and VI respectively.

2. RELATED WORK

In this section we review the wide range of research studies that have been done in analyzing student evaluations.

Authors Krishnaveni et al. [15] use basic text mining and Naïve Bayes classifier to classify the faculties into the respective classes (1 star – 5 star). They also introduce the concepts of student feedback weightage and ideal student to overcome the drawbacks encountered in the traditional feedback systems.

Authors Adesh et al. [16] perform sentimental analysis on the student feedback and classify the sentiments as positive, negative or neutral. They also perform comparative analysis using various machine learning techniques and conclude that multinomial naïve bayes classifier algorithm is more accurate when compared to support vector machine and random forest algorithms.

Authors Nasim et al. [17] use a hybrid approach for performing sentimental analysis on student feedbacks. They combine machine learning methods along with sentiment lexicons to analyze the sentiments expressed by students in their textual. They also perform comparative analysis and conclude that the best performing model was achieved using TF-IDF and domain-specific sentiment. However, their approach is limited to the computation of overall sentiment of the student feedback and does not include fine-grained analysis of student feedbacks.

Authors Altrabsheh et al. [18] discuss how feedback is collected via social media such as Twitter and they apply Sentiment Analysis to improve teaching called as Sentiment Analysis for Education (SA-E). They analyze the student feedback data collected from Twitter and perform pre-processing and feature extraction. They use Naïve Bayes and/or Support Vector Machine, which categorizes the whole post as either 'positive' or 'negative'.

Authors Aung and Myo [19] propose to analyze students text feedbacks using Lexicon based approach to predict the level of teaching performance automatically. To achieve this, they create a database of English Sentiment words as a lexical source to get polarity of the words. They then analyze the sentiment using intensifier words which are extracted from the feedback to get opinion result of teachers that can be strongly positive, moderately positive, weakly positive, strongly negative, moderately negative, weakly negative or neutral.

Authors Rajput et al. [20] apply text analytics methods on student's feedback data and obtain insights about teacher's performance with the help of tag clouds, and sentiment score. They suggest the computation of sentiment score to classify the feedback as either positive, negative, or neutral. They measure its performance, accuracy, recall, precision, and F-measure and find it to be very positive. They also demonstrate that the sentiment score is comparable to aggregated Likert scale-based score. They also show how word clouds visualization techniques can aid in getting unique insight into a teacher's performance which is typically not available via Likert-based scores.

In this paper we propose analyzing the qualitative end-of-course teacher evaluations with fine grained emotions such as 'anger', 'trust', 'sadness', 'joy', 'anticipation', 'fear', and 'disgust' with the help of National Research Council - NRC Emotion lexicon. We do this by combining the

word frequency and sentiment score to determine the overall sentiment or emotion associated with student feedback.

3. METHODOLOGY

This section contains the steps taken to process the student evaluation data. They are Data Collection, Data Extraction, Data Pre-Processing, Lexicon-Based Emotion Labeling and Visualization.

3.1. Data Collection

The dataset used for this work is collected from the Web-Based course evaluation system by UNC Charlotte. It is administered by a third-party Campus Labs. With the help of UNC Charlotte Center for Teaching and Learning, Campus Labs collect the student feedback for course evaluations. The student feedbacks for an instructor is collected for the terms of 2013 to 2020 including Fall, Spring and Summer sections of various courses handled by the instructor. We collect the html files from Campus Labs website for each of the semester. The data includes both quantitative and qualitative feedback. For this study we used qualitative feedback mainly focusing on Sentiment Analysis. Sample qualitative data shown in Table 1.

Table 1. Sample student feedback – qualitative

S.No	Comments
1	Instructor well prepared and shared good study materials.
2	The course work was good.
3	Flexible self-paced course.
4	Should be less group activities.
5	Best professor

3.2. Data Extraction

Once the data collection from Campus Labs is complete, jsoup [21] a Java library is used to process the html files and to extract the comments. The fields extracted from the html file are shown in Table 2. The data extracted consists of 959 records with the five attributes.

Table 2. Fields/Attributes Extracted from html file

S.No	Fields/Attributes
1	Year
2	Term
3	Course
4	Questions
5	Comments

3.3. Pre-processing

Pre-processing is very essential when dealing with textual data and it involves the removal of noisy and unwanted parts from the text. In this work we use the Python Natural Language Toolkit (NLTK) [22] to work with student evaluation data. The following steps are involved in pre-processing of student course evaluation comments: Tokenization, lower case, stop words removal.

3.3.1. Tokenization:

Tokenization is the process of breaking up a sentence into units (words). It also involves discarding punctuation marks from a sentence. An example is shown in Figure 1.

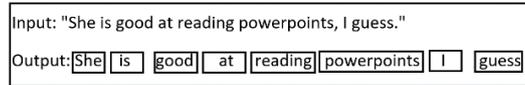


Figure 1. Tokenization

3.3.2. Changing to Lower Case:

The Student Feedback may contain both lower case and upper-case letters. For better performance it is essential for the text to be completely in lower-case letters, thus making it easier for the machines to process them.

3.3.3. Stop Words Removal:

Some words in a sentence do not contribute much to the meaning of the text but makes a sentence grammatically correct. For example, words like ‘is’, ‘was’, ‘it’, ‘are’ etc. These words are known as Stop Words. We use the list of stop words available in the Python Natural Language Toolkit (NLTK) [22] corpus to remove them from student feedback.

In the pre-processing step, certain invalid comments like ‘n/a’, ‘NA’, etc. are removed. The pre-processed dataset contains close to 1000 records in the dataset.

3.4. Emotion Labeling

Most important step in this process is Labeling the student feedback with different types of Emotion. We use the National Research Council - NRC Lexicon [9], [10] for this purpose. NRC Emotion lexicon consists of a list of English words and their association with eight basic emotions (anger, fear, anticipation, disgust, surprise, trust, joy, and sadness) and two polarities (positive and negative). The Annotations in the lexicon are at WORD-SENSE level. Each line has the format: <Term> <AffectCategory> <AssociationFlag> as shown in Figure 2.

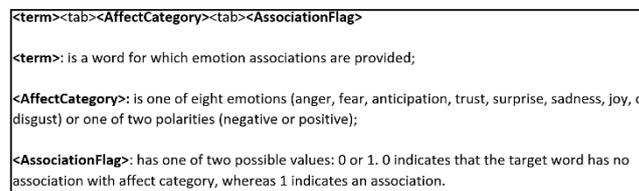


Figure 2. NRC emotion lexicon – word level annotation.

The Emotion Labeling steps are shown in Figure 3. Each of the student comments is processed and based on the Flag value in NRC lexicon if a match to word is found then the score is incremented accordingly. If a word is present multiple times, then the frequency score for that particular emotion will be incremented. After the entire comment is processed the Emotion which has the highest score is assigned as the class label for that student comment. If the final emotion score is zero, then those records are omitted from the dataset.

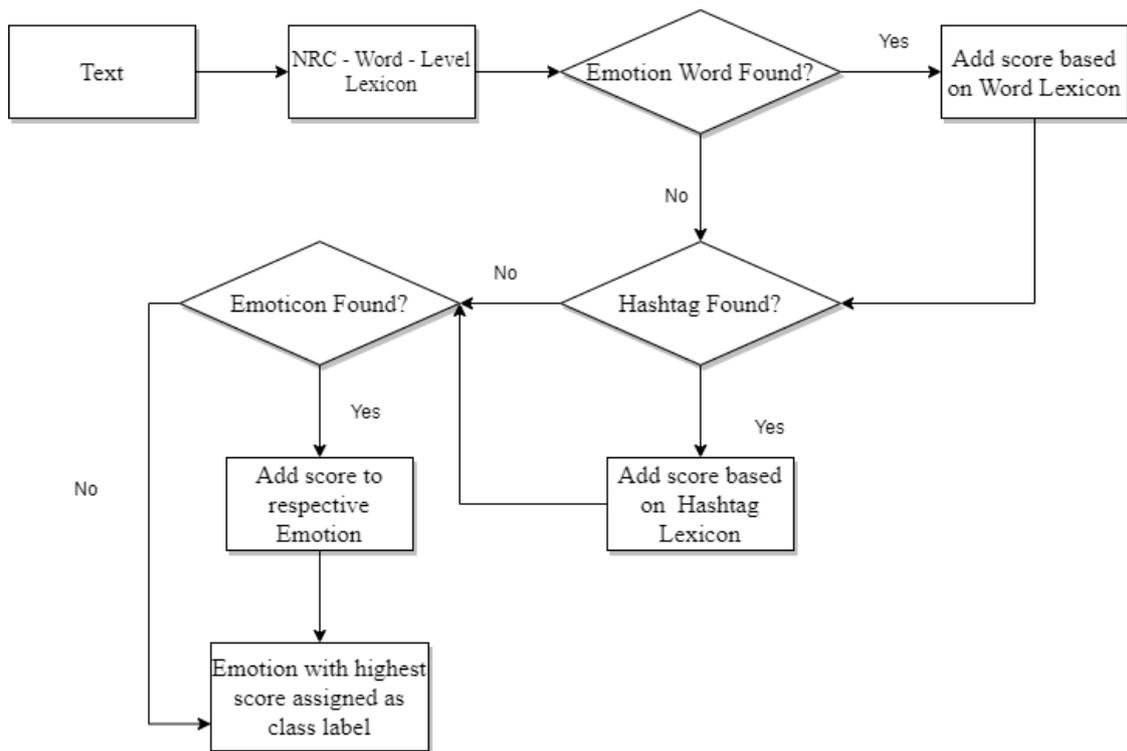


Figure 3. Emotion labeling process

3.5. Visualization:

Human mind is very visual [23] and data visualizations help in understanding a concept more clearly. The main focus of this paper is to identify emotions that the students feel toward courses delivered with changes including Light Weight teams, flipped classroom, and active learning methodologies. After labeling the students feedback with appropriate Emotion class, we visualize the results over the time period of 2013 to 2020 and results are analyzed. We use Tableau visualization software [24] for this purpose.

4. EXPERIMENTS AND RESULTS

This section contains information related to our experiments and the results obtained. The data for this study is collected from UNC Charlotte Campus Labs website. Sample qualitative data shown in Table 1. The data extracted has five attributes as shown in Table 2. For labeling the data - student feedback comments with different types of Emotion. We use the National Research Council - NRC Lexicon [10], [11].

In our work we perform two experiments. The first experiment includes basic emotions ‘anger’, ‘trust’, ‘fear’, ‘sadness’, ‘disgust’, ‘anticipation’, ‘surprise’, and ‘joy’. The second experiment includes polarities positive and negative along with the basic emotions.

4.1. Experiment 1 - Labeled with Basic Emotions:

In Experiment 1, the pre-processed data is passed to the system and it finds the word associated with 8 basic emotions for each of the student evaluation feedback response. The sentiment scores are then calculated based on the frequency of each of the emotion related words. The sentiment

with highest score is assigned as overall emotion of the student response. The results are shown on a temporal basis from 2013 until 2020 on the X-axis and the count of each emotion on the Y-axis in Figure 4. The results are almost same as the results of Emotion and Polarity graph plot, but without the two polarities ‘positive’ and ‘negative’. It is observed that emotion ‘trust’ has a growing trend through the time. We also see that ‘anticipation’ was high during the year 2014 which has gradually decreased in the year 2020.

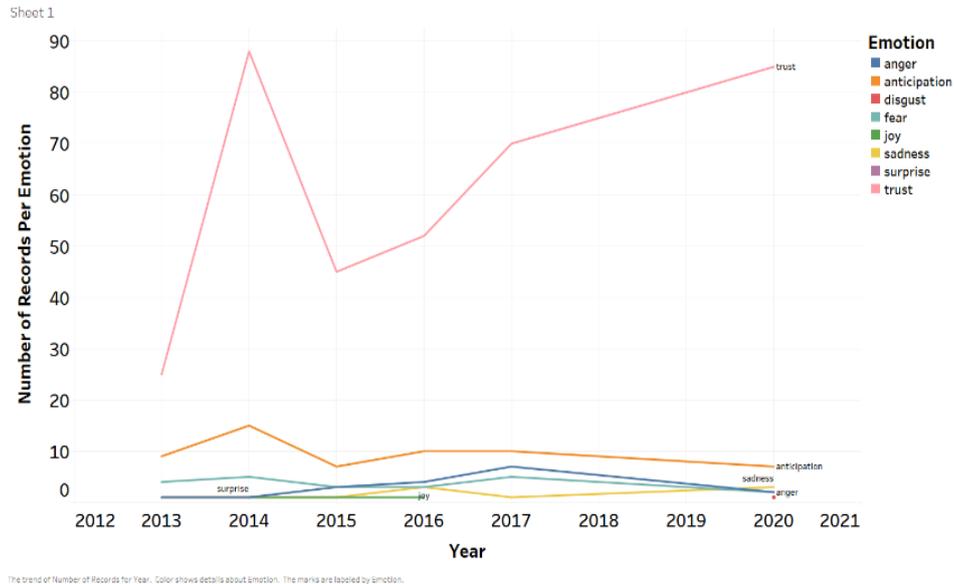


Figure 4. Experiment 1 – basic emotion

4.2. Experiment 2 – Labeled with Basic Emotion and Polarity:

In Experiment 2, the pre-processed data is passed to the system and it finds the word associated with 8 basic emotions and the polarity associated for each of the student feedback response. The sentiment scores are calculated based on the frequency of each emotion and polarity related words. The sentiment that has highest score is assigned as the overall emotion or polarity of the student response. The results are shown on a temporal basis from 2013 until 2020 on the X-axis and the count of each emotion on the Y-axis in Figure 5. It is observed that emotion ‘trust’ and polarity ‘positive’ has a growing trend through the time. We also observe that ‘anticipation’ was high during the year 2014 which has gradually decreased in the year 2020. We observe positive change in Emotion since year 2015 when active learning methodologies were introduced and implemented.

4.3. Sentiment Analysis and Emotion Detection in Student Evaluations – Word Cloud:

Word Cloud shows a visual representation of most frequently occurring words in a text and the size of the word in Word Cloud is proportional to its frequency in text. Word Cloud can be effectively used to solve text analysis tasks and evaluate it in a qualitative user study [25]. In this work we use the Word Cloud package in Python to create Word Clouds using the emotional words from the student evaluation data. During the emotion labeling step for each of the student feedback, the emotional words are recorded separately for each of the eight emotions and the positive and negative polarities. The words with negative emotions like ‘anger’, ‘fear’, ‘sadness’, ‘disgust’, and ‘negative’ polarity are taken as negative word list from the NRC Emotion Lexicon



Figure 10. WordCloud-2017. Most frequent word appears with largest font. Negative words in red. Positive words in green.

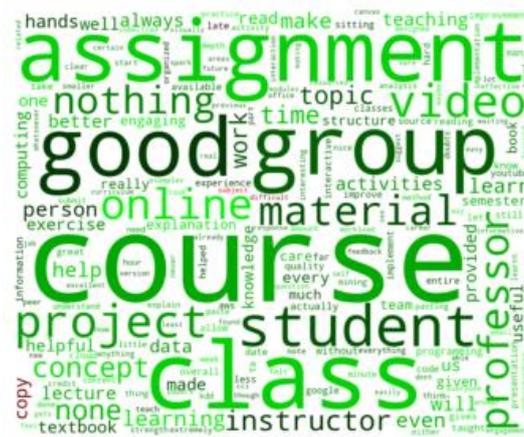


Figure 11. WordCloud-2020. Most frequent word appears with largest font. Negative words in red. Positive words in green.

5. CONCLUSIONS

In this work, we mine and detect emotions from the qualitative feedbacks given by the students in the end of semester course evaluations. We find two sentiment polarities 'positive' and 'negative' and also identify eight basic human emotions like 'joy', 'surprise', 'trust', 'anticipation', 'disgust', 'anger', 'fear' and 'sadness' from the feedback response. We evaluate the impact of Active Learning methodologies like Flipped Classroom approach [14] and Light Weight Teams [12], [13] implemented in classrooms during the time period of 2015 to 2020. We find that the words associated with positive sentiments and the emotion 'trust' have increased during these years. We also find that the words associated with negative sentiments and emotions have decreased during these years. Hence, we conclude that the implementation of Active Learning methodologies like Flipped Classroom approach and Light Weight teams have positive impact on students and their learning experience. In our future work we plan to focus on diverse groups in computing discipline.

REFERENCES

- [1] Darwin S. (2016). *Student Evaluation in Higher Education: Reconceptualising the Student Voice*. Cham: Springer International Publishing AG.
- [2] Flood Page, C. (1974). *Student evaluation of teaching: The American experience*. London: Society for Research Into Higher Education.
- [3] Kulik, J. (2001). Student ratings: Validity, utility and controversy. *New directions for institutional research*, 109.
- [4] Marsh, H. W. (1987). Students' evaluations of university teaching: research findings, methodological issues and directions for future research. *International Journal of Educational Research*, Vol 11, pp. 253–388.
- [5] Dictionary, C., Cambridge Dictionary, On-line at <https://dictionary.cambridge.org>.
- [6] Spooren, P., Brockx, B., & Mortelmans, D. (2013). On the validity of student evaluation of teaching: The state of the art. *Review of Educational Research*, Vol. 83, No. 4, pp.598-642.
- [7] F. A. Pozzi, E. Fersini, E. Messina, and B. Liu, *Sentiment analysis in social networks*. Morgan Kaufmann, 2016.
- [8] "Emotion." Merriam-Webster.com Dictionary, Merriam-Webster, <https://www.merriam-webster.com/dictionary/emotion>. Accessed 19 Dec. 2020.
- [9] Ortony, A., & Turner, T. J. (1990). What's basic about basic emotions?. *Psychological review*, Vol. 97, No. 3, pp. 315.
- [10] Mohammad, S. M., & Turney, P. D. (2013). Crowdsourcing a word–emotion association lexicon. *Computational Intelligence*, Vol. 29, No. 3, pp. 436-465.
- [11] Mohammad, S. M., & Turney, P. D. (2010, June). Emotions evoked by common words and phrases: Using Mechanical Turk to create an emotion lexicon. In *Proceedings of the NAACL HLT 2010 workshop on computational approaches to analysis and generation of emotion in text*, Association for Computational Linguistics, pp. 26-34.
- [12] Latulipe, C., Long, N. B., & Seminario, C. E. (2015). Structuring flipped classes with lightweight teams and gamification. In *Proceedings of the 46th ACM Technical Symposium on Computer Science Education*, pp. 392-397.
- [13] MacNeil, S., Latulipe, C., Long, B., & Yadav, A. (2016). Exploring lightweight teams in a distributed learning environment. In *Proceedings of the 47th ACM Technical Symposium on Computing Science Education*, pp. 193-198.
- [14] Maher, M. L., Latulipe, C., Lipford, H., & Rorrer, A. (2015). Flipped classroom strategies for CS education. In *Proceedings of the 46th ACM Technical Symposium on Computer Science Education*, pp. 218-223.
- [15] Krishnaveni K.S., Pai, Rohit R., Iyer, Vignesh, "Faculty Rating System Based on Student Feedbacks Using Sentimental Analysis." (2017). In *International Conference on Advances in Computing, Communications, and Informatics*. IEEE, 2017. 1648–1653. Web.
- [16] Adesh, N. D., Dsouza, D. D., Deepika, Nayak, D. P., Machado, E. J., (2019, June). Sentimental Analysis of Student Feedback using Machine Learning Techniques. In *International Journal of Recent Technology and Engineering*, ISSN: 2277-3878, Volume-8, Issue-1S4.
- [17] Z. Nasim, Q. Rajput and S. Haider, (2017). "Sentiment analysis of student feedback using machine learning and lexicon based approaches," *International Conference on Research and Innovation in Information Systems*, Langkawi, 2017, pp. 1-6.
- [18] Altrabsheh, N., Gaber, M., & Cocea, M. (2013, June). SA-E: sentiment analysis for education. In *International Conference on Intelligent Decision Technologies*, Vol. 255, pp. 353-362.
- [19] Aung K. Z. & Myo N. N., (2017) . "Sentiment analysis of students' comment using lexicon based approach," In *IEEE/ACIS 16th International Conference on Computer and Information Science (ICIS)*, Wuhan, 2017, pp. 149-154, doi: 10.1109/ICIS.2017.7959985.
- [20] Rajput, Q., Haider, S., & Ghani, S. (2016). Lexicon-Based Sentiment Analysis of Teachers' Evaluation. *Applied Computational Intelligence and Soft Computing*, pp. 1.
- [21] Hedley, J. (2009). jsoup: Java html parser. 2009-11-29 [2015-06-12] <http://jsoup.org>.
- [22] Bird, S., Klein, E., & Loper, E. (2009). *Natural language processing with Python: analyzing text with the natural language toolkit*. " O'Reilly Media, Inc."
- [23] Aparicio M., & Carlos J. Costa. (2015). Data visualization. *Commun. Des. Q. Rev*, pp. 7–11. DOI:<https://doi.org/10.1145/2721882.2721883>.

- [24] Hanrahan, P. (2003). Tableau software white paper-visual thinking for business intelligence. Tableau Software, Seattle, WA.
- [25] Heimerl, F., Lohmann, S., Lange, S., & Ertl, T. (2014). Word cloud explorer: Text analytics based on word clouds. In System Sciences (HICSS), 2014 47th Hawaii International Conference on (pp. 1833-1842). IEEE.