2020 IEEE Region 10 Symposium (TENSYMP), 5-7 June 2020, Dhaka, Bangladesh

Understanding Customer Sentiment: Lexical Analysis of Restaurant Reviews

Jinat Ara[1], Md. Toufique Hasan[2], Abdullah Al Omar[3], Hanif Bhuiyan[4,5]
[1]*Jahangirnagar University, Dept. of CSE, Savar, Dhaka, Bangladesh*
[2]*European University of Bangladesh, Dept. of CSE, Dhaka, Bangladesh*
[3]*University of Asia Pacific, Dept. of CSE, Dhaka, Bangladesh*
[4]*Data61, CSIRO*, [5]*Queensland University of Technology, CARRS-Q, Queensland, Australia*
aracse2014@gmail.com, toufique.hasan@hotmail.com, omar.cs@uap-bd.edu
Corresponding author's E-mail: hanif.bhuiyan@data61.csiro.au

***Abstract*—Understanding customer's sentiment (satisfaction or dissatisfaction) is considered as valuable information for both the potential customers and restaurant authority. However, analyzing customer reviews (unstructured texts) one by one is a difficult task and also practically impossible when the number of reviews is enormous. Therefore, it seems conceivable to have a mechanism to analyze customer reviews automatically and provide the necessary information in a precise way. Here, we introduce a Natural Language Processing (NLP) based opinion mining methodology to analyze the customer opinion automatically. For that, first, a captive portal is used to collect customer's reviews. Then, the opinion mining technique is applied to analyze the reviews to explore customer sentiment about food quality, service, environment, etc. A data-driven experiment is conducted to evaluate the proposed methodology. The experiment result showed the effectiveness of the proposed method for retrieving and analyzing customer sentiment.**

***Index Terms*—customer opinion, lexical analysis, opinion mining, sentiment analysis, captive portal.**

I. Introduction

Customers are the lifeblood of the restaurant. Maintaining customer satisfaction level is essential to keep the restaurant business well. Therefore, restaurant authority always tries to understand customer perception through several processes [1] [2]. Among them, one of the widely used methods is the review system. Restaurant authority allows their customer to review their services based on several parameters (such as food quality, management, etc.) [3] [5]. Usually, these reviews are "star rating review" (figure1. a) [1] [4]. However, most of the time this star rating review does not reflect the actual feed of customer. Customers might give the same star rating for the entire question or may provide star rating without reading questions appropriately, which may provide the wrong information to restaurant authority. To solve these issues, restaurant authority has started online written review (figure1. b) to know the actual feedback of customer [6] [7]. The online written review has taken through the restaurant web portal where some pre-defined questions are asked to the customer to express their opinions. This instant review helps customers to express their opinions more interactively [8] [9].

Day by day online written reviews are widely adopted by restaurants to understand the customer sentiments implicitly. Understanding customer sentiment is one of the significant factors for the improvement of restaurant quality in terms of food, management, etc. However, analyzing the one by one large number of reviews is a challenging task. Therefore, an automatic process is highly recommended to analyze the review to understand customer's perception. The goal of this paper is to develop an automated review process that analyzes the customer's written reviews, calculates the strength, and predicting the customer's preferences with the restaurant quality.

In general, the customer's written reviews are in natural

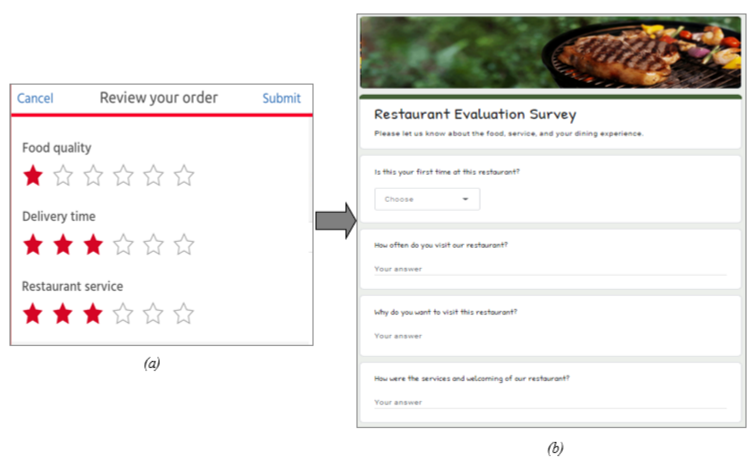

Fig. 1. Web portal view for star rating customer review (a) and written format customer review (b) process.

language format includes unstructured paradigm that difficult to handle. Several research has been proposed to process this unstructured data. Among them, several research works showed significant improvement and proved that Natural Language Processing technique (NLP) is an effective solution to analyze the unstructured data [10] [16]. Sentiment analysis using NLP is also an effective solution to analyze these unstructured data or opinions [11]. Sentiment analysis (also known as opinion mining) refers as lexical analysis that useful for opinion mining. It also helps to judge a review strength that automatically calculates the sentiment polarity (pos,neg, neutral) and identify whether review expresses the positive opinion or negative opinion or neutral opinion. Therefore, this work is focused on developing an instant written review system that performs an NLP based sentiment analysis method based on customer's opinions. This analysis aims to make customer's opinions useful for the restaurant authority to predict customers preferences and restaurant-quality.

The reminder of the paper is organized as follows. At first, section 2 provides information about background study on the area of customer opinion analysis. Then, Section 3 provides details strategy of the opinion mining process for customer opinion. Next, a compressive experiment result shows in section 4. Finally, conclude the proposed work through the conclusion in section 5.

II. Related Work

Recently, a significant number of works proposed for automatic summarization of the restaurant's customer opinion [5]. They addressed that lexical analysis of customer's opinions is effective to represent positive, negative and neutral strength simultaneously [12] [13]. Natalia et al. [9] proposed a lexical analysis based opinion mining approach to predict food item's popularity. Chinsha et al. [14] proposed a sentiment analysis process for classifying several features from customer reviews. These two proposed approaches are effective for the lexical analysis of the customer's opinion. Recently, researchers are interested in analyzing customer reviews through machine learning techniques [11] [15]. Akshay et al. [17] adopt a machine learning algorithm to classify the customer's perception regarding positive, negative, or neutral ratings. Hu et al. [6] proposed a sentiment analysis process for frequent feature extraction through the bootstrapping technique. These two approaches are effective for rating review process only. Besides, we consider the customer's written reviews to understand user satisfaction or dissatisfaction level. It might be more helpful for the restaurant authority to improve their restaurant quality.

III. Methodology

In this section, we describe the proposed NLP-based opinion mining technique on customer reviews or opinions for understating the customer's sentiment. The proposed work describes through configuring networking devices and captive portal installation. Next, lexical analysis is performed on the customer's reviews or opinions and evaluates the effectiveness of the proposed methodology. In detail, the working methodology is described below.

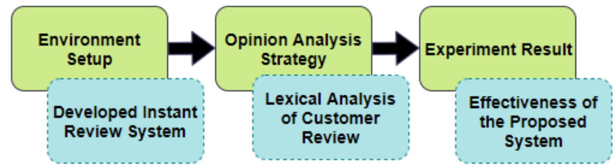

Fig. 2. System workflow

A. Environment Setup

In this section, we develop an instant review system where customers can express their opinion immediately. The instant review system is developed into two steps: networking device configuration and captive portal implementation. At first, we configure a networking device or router to implement the captive portal into the networking device. After the captive portal configuration, customers can directly connect with the restaurant web portal based on network availability. The environment setup process describes in bellow.

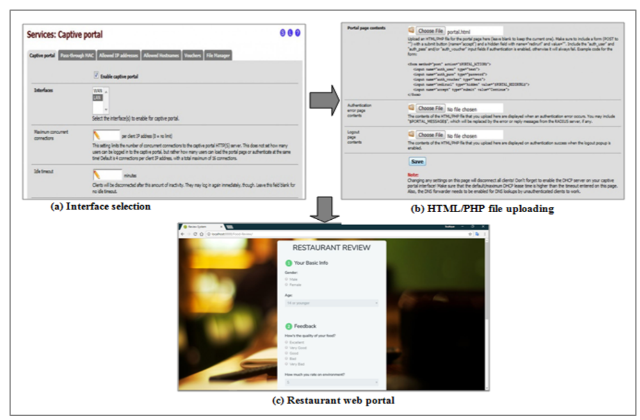

Fig. 3. Environment setup process (captive portal service).

1) Networking device configuration: A networking device (router) is configured to provide support during the captive portal installation process. It provides the communication facilities between communication devices (mobile/laptop/iPad) and restaurant web portal. After the configuration process, the captive portal (next section) allows the user to interact with the system and re-direct to the restaurant web portal.

2) Captive portal implementation: To perform the customer's opinion collection process captive portal is implemented (figure3. a) in the networking device for introducing a web page. The captive portal defines as a web page that works in a public-access networking environment. After configuring the captive portal, an HTML/PHP file is uploaded (figure3. b) which known as a landing page that connects the user with the restaurant web portal. Then, a communication device (mobile/laptop/iPad) can connect with the available network and gets a popup window with questionnaires (figure3. c).

B. Opinion Analysis Strategy

This section describes the NLP based opinion analysis process on customer's reviews. The proposed work has completed into five steps, as shown in figure 4: Opinion gathering, Data collection & preprocessing, Word extraction, Opinion mining and Opinion classification. At first, the opinion gathering

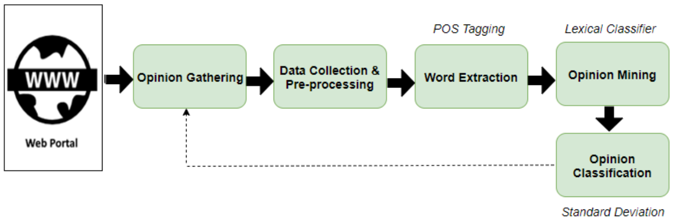

Fig. 4. Lexical analysis for opinion mining of customer reviews.

process is performed through a review process where users or customers of the restaurant can directly interact with the reviewing process and express their opinion regarding the predefined questions. Then, the data collection preprocessing step performs through extracting reviews or opinions. Next, the word extraction process performs on process data to extract the most valuable words. After that, the opinion mining performs through the SentiStrength classifier or lexical classifier [6] [18] to calculate the value of opinion expressed words to understand the user's interest or opinion about restaurant quality (service, food quality, environment etc.). Finally, opinion classification performs through the standard deviation technique to summaries the value of customer opinion to predict positive, neutral and negative opinions. The opinion analysis process describes in detail bellow.

1) Opinion gathering: The goal of this section is to gather the user's opinions. The opinion gathering process performs through the restaurant web portal (environment setup section) where users can directly interact and write their opinions through digital devices (mobile, laptop, tablet, etc.). For the opinion gathering process, we asked the customers to express their opinions about restaurant quality. After gathering reviews, automatically collected and stored data into the database server. The review generation process shows in figure 5.

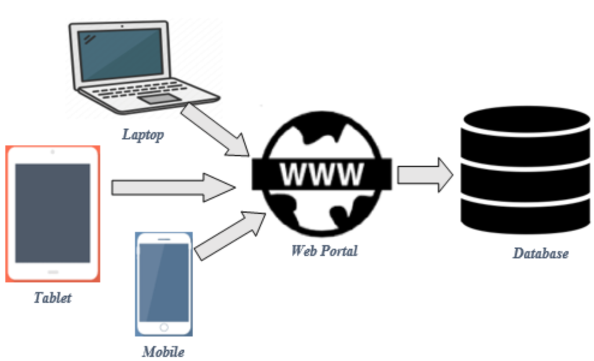

Fig. 5. Customer opinion gathering process

2) Data collection & preprocessing: In this section, a data collection module extracts data from the opinion database and performs a preprocessing step on extracted data. Preprocessing step performs to make data structured as customer opinions are unstructured with various annotations. To classify the unstructured data, apply stop word for eliminating entire stop words and make dataset1. After generating dataset1, eliminate all the linking words and all the punctuation's ("."), space ("-"), commas (","), semicolon (";"), hash ("-"), etc. to make dataset2.

3) Word extraction: In the data collection & preprocessing stage, dataset2 is made through structured data or opinion that free from all notions. In this section, we extract the most valuable opinion words from dataset2 that carries positive and negative value. For extracting the opinion words, part-of-speech is applied on dataset2 that finds the noun, pronoun, adjective, verb, adverb, etc. It extracts all the parts of speech words and makes dataset3. POS tagging is essential for any lexical analysis because it finds the most valuable or sentiment words that are responsible for predicting positive or negative or neutral strength of the opinion.

4) Opinion mining: In this section, SentiStrength is applied to dataset3 and calculates the strength of extracted words to predict customer opinion. SentiStrength is a lexicon classifier which calculates opinion strength regarding positive or negative value by annotating 1 (positive) to 5 (extremely positive) and -1 (negative) to -5 (extremely negative). Its analysis the

TABLE I
POSITIVE AND NEGATIVE STRENGTH OF CUSTOMER OPINION WORD.

Its decorations are nice & the taste is good.			Services were fine, welcoming and was comparatively hospitable.		
Extracted Words	**Pos**	**Neg**	**Extracted Words**	**Pos**	**Neg**
Decoration	1	-1	Service	1	-1
nice	3	-1	fine	3	-1
taste	1	-1	welcoming	2	-1
good	3	-1	comparatively	1	-1
-	-	-	hospitable	1	-1

customer's opinions considering positive words and negative words where positive word expressed through ≥ 1 score and negative word expressed through <1 score. From the opinion analysis, it could possible to retrieve the customer's opinion whether it is positive or negative. For example, table 1 shows two individual customer opinions "Its decorations are nice & the taste is good" and "Services were fine, welcoming and was comparatively hospitable". From its SentiStrength score, it seems that a particular opinion consists of both positive scores and negative scores simultaneously.

5) Opinion classification: The opinion mining section shows that the customer's opinion represents both positive and negative strength simultaneously. So, it's relatively difficult to conclude about any review, whether it is a positive or negative opinion. For predicting the exact strength of the opinion, the Standard Deviation (SD) technique is applied. Standard Deviation (SD) is a static measurement technique that concludes the overall strength of a particular review through SD value. Higher SD value for positive opinion word set represents higher positive strength and lower SD value for negative opinion word set represents lower negative strength respectively. However, the opinion's rating score influences the SD value where one higher rating score (2 to 5) could responsible for increasing and one lower rating score (-5 to 1) could responsible for decreasing the SD value. For instance, table II shows a sample of data set with their mean and SD value. In positive dataset, higher rating of dataset1 is 3 and 4 for dataset2 respectively. Among these two datasets, sample dataset2 has higher SD (0.862) for its higher rating score (4) which represents the maximum distance between the sample data and the mean. Similarly in the negative dataset, -2 and -3 is the higher rating score for sample dataset1 and sample dataset2 respectively. Sample dataset2 has higher SD (0.616) than sample dataset1 (0.445) as it has the higher rating score (-3). According to the SD value, a higher SD value of positive class with a lower SD of the negative class is more effective for classifying the positive strength of customer opinion. Figure6 shows both positive and negative strength of five customer's opinions. In figure6, opinion1, Opinion2, Opinion3, and Opinion5 have higher positive strengths (1.548, 1.680, 1.794, and 1.376) than the positive strength of opinion4 (0.577). It seems that, among five reviews, four reviews (1, 2, 3, and 5) are positively classified and another one (4) is negatively classified.

TABLE II
CLASSIFICATION OF SD VALUE CONSIDERING PROPOSED APPROACH

Positive Dataset		Negative Dataset	
Sample Dataset1	**Sample Dataset2**	**Sample Dataset1**	**Sample Dataset2**
1	1	-1	-1
1	1	-1	-1
1	1	-1	-1
1	1	-2	-1
1	1	-1	-1
1	1	-1	-1
1	1	-1	-1
1	1	-1	-1
3	4	-1	-3
2	1	-2	-2
1	1	-2	-1
Sum=14	**Sum=14**	**Sum=14**	**Sum=14**
Mean=1.272	**Mean=1.272**	**Mean=1.272**	**Mean=1.272**
SD=0.616	**SD=0.862**	**SD=0.445**	**SD=0.616**
Lower Positive	Higher Positive	Lower Negative	Higher Negative

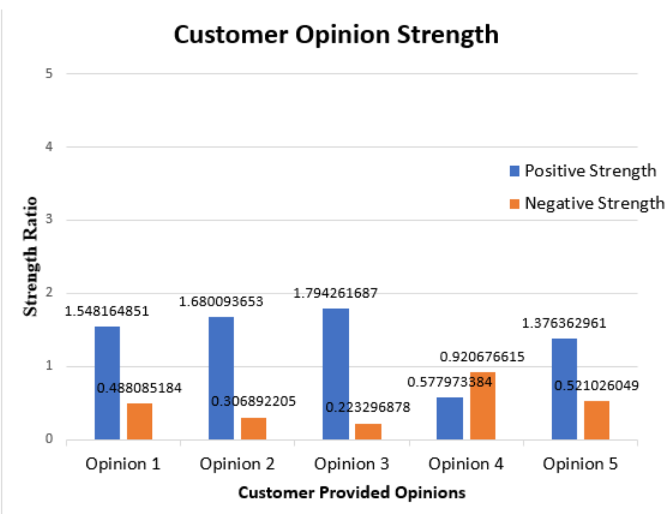

Fig. 6. Strength of five customer's opinions (positive/negative).

IV. EXPERIMENT

In this section, we present the experimental result to evaluate the proposed opinion mining approach. For experimenting, we invited almost 35 participants/customers to provide their valuable opinions through the restaurant web portal. They expressed their opinion for almost twenty questions including about restaurant quality, food quality, services, environment, welcoming, price, online food order service, suggestions, etc. The experiment was conducted on 700 opinions that were collected from the restaurant web portal.

The experiment results evaluated through the SD value of both positive and negative opinion word sets. To evaluate the SD value, manual inspection was performed on the experiment result. 10 volunteers were recruited to justify the result through the proposed process. Volunteer checked both positive and negative strength sets. Their result suggested that individual strength set (positive & negative) having 0.75 SD is perfect to predict the positive and negative opinions. They also depict that $\geq$ 0.75 SD value represents the highly positive opinion and <0.75 SD value represents the highly negative opinion. Therefore threshold value was set 0.75. For evaluating the customer's opinion, this threshold value was effective. From the classification result and our observation, it revealed that higher SD value for positive opinion word set rated as a positive opinion and higher SD value for negative opinion word set rated as negative opinion.

After classifying the SD of individual opinion word sets, a comparison process performed on both positive and negative opinion word set together. The comparison result showed that opinion rated as positive when SD of positive opinion word set is higher than negative opinion word set. Similarly, opinion rated as negative when SD of negative opinion word set is higher than positive opinion word set. In contrast, similar SD of both positive and negative opinion word set refers as neutral opinion (not rated). According to this comparison process we analyze and classify our experimental result. The experiment result shows that (table2) among 35 customer's opinions, only 20 customer's opinions have $\geq$0.75 SD and marked as positive opinion, 5 customer's opinions have similar SD for both positive and negative strength set that classified as neutral (not rated) and 10 customer's opinions have <0.75 SD value that classified as negative opinion.

To evaluate the effectiveness of the proposed approach, manual observation performs on extracted customer's opinions. For conducting this observation, 10 volunteers were selected. Our volunteers were manually observing these opinions and retrieve 300 opinions as positive and 225 opinions as negative. Among 700 opinions, manual observation could not conclude 175 opinions whether the opinions are positive or

TABLE III
EXPERIMENT RESULT OF CUSTOMER OPINION ANALYSIS.

Proposed Approach			
Total Number of Customer (20 questions)	Number of customer (Positive Opinion)	Number of customer (Neutral Opinion)	Number of customer (Negative Opinion)
35	**20**	**5**	**10**
Number of Opinion	Number of Positive Opinion	Number of Neutral Opinion	Number of Negative Opinion
700	**400**	**100**	**200**
Total Accuracy (%) = 85.714 %			
Manual Observation			
Total Number of Customer (20 questions)	Number of Positive Opinion	Number of Neutral Opinion	Number of Negative Opinion
700	**300**	**175**	**225**
Total Accuracy (%) = 75.0 %			

negative. In contrast, table2 shows that the proposed approach identify 400 opinions as positive and 200 opinions as negative. Among 700 opinions, the proposed approach could not conclude the strength of 100 opinions whether the opinions are positive or negative. From both manual observation and experiment results, it concludes that the proposed approach is effective to identify the positive and negative opinion strength more accurately than manual observation. That seems that manual observation predicts more inaccurate result than an automatic process. From this perception, the error rate of the proposed approach is 14.285% which indicates the accuracy of the proposed approach is 85.714%. In contrast, the error rate of the manual observation is 25% that shows that the accuracy of the manual observation is 75.0%.

However, accuracy of the proposed approach (85.714%) shows the effectiveness of the proposed approach. It can concludes that customer opinions are useful to evaluate restaurant progress status. This analysis helps to understand customer preferences, restaurant quality, and overall strategy.

V. CONCLUSION

In this paper, an NLP-based opinion mining process has developed for the analysis of the restaurant's customer opinion to predict about restaurant quality and future improvement. The opinion mining performs through the SentiStrength classifier to reveal the positive, negative and neutral strength of customer opinions. This analysis performs on 700 opinions provided by the customer. This approach reveals that customer opinions are effective in analyzing customer perception of restaurant quality. From the experimental result and manual analysis, it concludes that the proposed approach and manual observation are both effective for predicting the customer's opinion. The experiment result shows that the efficiency of the proposed opinion mining approach is 85.714% to retrieving customer's opinions. It provides future direction to analyze more customer's opinions in order to classify the customer's preferences more correctly.

REFERENCES

[1] Y.M. Aye. and S.S. Aung, Senti-Lexicon and Analysis for Restaurant Reviews of Myanmar Text. International Journal of Advanced Engineering, Management and Science, 2018. 4(5).

[2] H. Saeed, Near-field communication sensors and cloud-based smart restaurant management system. in 2016 IEEE 3rd World Forum on Internet of Things (WF-IoT). 2016. IEEE.

[3] M. Nakayama. and Y. Wan, The cultural impact on social commerce: A sentiment analysis on Yelp ethnic restaurant reviews. Information Management, 2019. 56(2): p. 271-279.

[4] M. Lee. and R. Grafe, Multiclass sentiment analysis with restaurant reviews. Final Projects from CS N, 2010. 224.

[5] G. Somprasertsri. and P. Lalitrojwong, Mining Feature-Opinion in Online Customer Reviews for Opinion Summarization. J. UCS, 2010. 16(6): p. 938-955.

[6] M. Hu. and B. Liu, Mining and summarizing customer reviews. in Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining. 2004. ACM.

[7] Y. Lee, Building Place-Specific Sentiment Lexicon. in Proceedings of the 2nd International Conference on Digital Signal Processing. 2018. ACM.

[8] J. Guerreiro. and P. Rita, How to predict explicit recommendations in online reviews using text mining and sentiment analysis. Journal of Hospitality and Tourism Management, 2019.

[9] H. Bhuiyan. Oh. Kyeong-Jin. Hong. Myung-Duk, and Jo. Geun-Sik, An effective approach to generate Wikipedia infobox of movie domain using semi-structured data,Journal of Internet Computing and Services. 2017. volume: 18, number: 3, p. 49–61.

[10] N. Klyueva, Food-Related Sentiment Analysis for Cantonese. in Proceedings of the 32nd Pacific Asia Conference on Language, Information and Computation: 25th Joint Workshop on Linguistics and Language Processing. 2018.

[11] D. Lee, O.-R. Jeong, and S.-g. Lee. Opinion mining of customer feedback data on the web. in Proceedings of the 2nd international conference on Ubiquitous information management and communication. 2008. ACM.

[12] Y.M. Aye. and S.S. Aun, Enhanced Sentiment Classification for Informal Myanmar Text of Restaurant Reviews. in 2018 IEEE 16th International Conference on Software Engineering Research, Management and Applications (SERA). 2018. IEEE.

[13] Y. Fang, H. Tan, and J. Zhang, Multi-strategy sentiment analysis of consumer reviews based on semantic fuzziness. IEEE Access, 2018. 6: p. 20625-20631.

[14] T. Chinsha. and S. Joseph, Aspect based opinion mining from restaurant reviews. International Journal of Computer Applications, 2014. 975: p. 8887.

[15] P.Ray. and A. Chakrabarti, A Mixed approach of Deep Learning method and Rule-Based method to improve Aspect Level Sentiment Analysis. Applied Computing and Informatics, 2019.

[16] H. Bhuiyan, J. Ara, R. Bardhan, M. R. Islam: (2017,September). Retrieving youtube video by sentiment analysis on user comment. In 2017, IEEE International Conference on Signal and Image Processing Application (ICSIPA), (pp. 474-478). IEEE

[17] A. Krishna, Sentiment Analysis of Restaurant Reviews Using Machine Learning Techniques, in Emerging Research in Electronics, Computer Science and Technology. 2019, Springer. p. 687-696.

[18] J. Ara. and H. Bhuiyan, Semantic Recognition of Web Structure to Retrieve Relevant Documents from Google by Formulating Index Term. In 2019, Computer Communication and Computational Sciences. Advances in Intelligent Systems and Computing, vol 924. Springer, Singapore.