American Journal of Embedded Systems and Applications
2022; 9(2): 66-72
http://www.sciencepublishinggroup.com/j/ajesa
doi: 10.11648/j.ajesa.20220902.12
ISSN: 2376-6069 (Print); ISSN: 2376-6085 (Online)

Sentiment Mining and Aspect Based Summarization of Opinionated Afaan Oromoo News Text

Wegderes Tariku[1], Million Meshesha[2], Ashebir Hunegnaw[3], Kedir Lemma[4]

[1]Department of Information Systems, Mizan Tepi University, Tepi, Ethiopia

[2]Department of Information Science Addis Ababa University, Addis Ababa, Ethiopia

[3]Department of Management Information Systems, Mettu University, Mettu, Ethiopia

[4]Deprtment of Information Technology Ambo University, Ambo, Ethiopia

Email address:

wegdreset@gmail.com (Wegderes Tariku), kedirnaw1999@gmail.com (Kedir Lemma)

To cite this article:

Wegderes Tariku, Million Meshesha, Ashebir Hunegnaw, Kedir Lemma. Sentiment Mining and Aspect Based Summarization of Opinionated Afaan Oromoo News Text. *American Journal of Embedded Systems and Applications*. Vol. 9, No. 2, 2022, pp. 66-72. doi: 10.11648/j.ajesa.20220902.12

Received: August 17, 2022; **Accepted**: September 7, 2022; **Published**: September 19, 2022

Abstract: Studying the specific subject of opinion mining has been a popular research area as a means of overcoming the challenge of user-generated content on the web, which can be challenging to manually collect, comprehend, summarize, and analyze for decision-making. Even though there are three various levels at which opinion mining can be done, the detail and complexity of feature level opinion mining outweighs its disadvantages. The goal of this research is to provide sentiment mining and aspect-based opinion summaries of service reviews in Afaan Oromo for Oromia Radio and Television Organization (ORTO). 400 reviews in all were gathered and used for news-related purposes from ORTO. The model has five elements, including document inspection, pre-processing, aspect extraction, polarity detection, and aspect-based sentiment summary, as well as a bar chart to show aspect-based sentiment summation. Five different processes make up the model: document review, pre-processing, aspect extraction, polarity detection, and aspect-based sentiment summarization. A bar chart is also utilized to visually depict aspect-based opinion polarity. For positive classes, 90% precision and 87% recall are accomplished, while for negative classes, 87% precision and 89.7% recall are attained. The main issue identified in this study is that users tend to express their opinions in a context-based or indirect manner. They could express their negative feelings with pleasant words or the opposite. Therefore, more research is required before the algorithm will take context-based or semantic opinion mining into account.

Keywords: Opinionated Afaan Oromo News Texts, Aspect Level Sentiment Mining, Sentiment Summarization, Lexical Database, Oromia Radio and Television Organization

1. Introduction

User-generated content (UGC) is expanding at a rapid rate according to the founding principles of web 2.0 [1]. Online on the web, a large amount of user-generated material can be easily generated in several domains. An opinion holder from a variety of review sites, blogs, discussion forums, social media review sites, etc. may put this UGC online. When making decisions, people frequently employ feelings, attitudes, opinions, or emotions that can be stated online against a certain entity to determine "what is the opinion of the opinion holder on some particular entity (i.e. object)?" [1]

Opinion mining or sentiment mining is the practice of determining the attitude of the person who holds an opinion toward anything [1]. As a result, sentiment mining addresses the issue of waiting for a response from friends [3] in several fields. It aids the company in enhancing its goods and services. Even though it is often regarded as a data mining technique, opinion mining is currently a popular research area in NLP [4]. Three granularity levels can be used to categorize sentiment mining [2]: Sentiment analysis at the sentence level is used to categorize review sentences into two concepts: subjective and objective. Positive or negative sentiment orientation is one way to categorize subjective.

The objective ones, however, are facts [5]. The total sentiment expressed by an opinion holder toward a specific entity is determined using a document level sentiment mining task. There is no method to determine opinion orientation on each and every specific aspect at the document level. All it is doing is giving the review document a basic sentiment direction [1]. It is impossible to pinpoint the precise elements that an opinion bearer loves and despises at the sentence or document level [1]. Since aspect level opinion mining deals with the elements or qualities of an object (such as a product, movie, etc.) at a finer grained level of analysis, this is possible [2]. Because of this, it is possible to determine an object's opinion at the aspect level one level of detail.

Opinion mining is done using different methods and techniques. In this study, we used rule-based method to detect the aspect of the opinions and to predict the polarity of the sentiments and then summarize the polarity of opinions based on each aspect using bar chat. In this study, service review for ORTO is considered; because it is one of the most popular organization broadcasting news, drama, music, sport programs and other educational services. Different scholars on sentiment mining, and summarization are being undertaken and done for languages such as English [6], French, Turkey [7], Arabic [8] and Amharic [9-11]. Summarization at aspect level in Afaan Oromo language that show the product or service features that are liked or disliked by the users or the customers never attempted until now, even though the volume of Afaan Oromoo text on the web is increasing rapidly. Therefore, this study aims to develop sentiment mining and aspect-based opinion summarization of service review for Oromia ORTO. To this end, this study attempts to address and answer the followings research questions.

1. How Data (corpus) can be prepared for service review in Afaan Oromo language?
2. How to develop suitable prototype for polarity classification?
3. How to summarize aspect level opinion in service review domain?
4. To what extent the prototype performs opinion prediction and aspect level opinion summarization?

2. Related Works

Tullu Tilahun [9] conducted feature level model for Amharic language opinion mining in three different domains by employing manually crafted rules and lexicon. Two investigates for feature extraction and sentiment word determination using 484 opinions was tasted and gained average precision of 95.2% and 26.1% recall for feature extraction, and average precision of 78.1% and 66.8% recall for determination of opinion words.

Selama Gebremeskel [11] on the other hand considered sentiment mining for opinionated Amharic text at using domain specific and general opinion terms into negative and positive including, contextual valence shifters by using a total of 303 from movie and newspaper reviews. The lexical of Amharic opinion terms were constructed for assigning initial polarity value. System prototype was developed fot authenticatio of the proposed model and the algorithms designed. Three experiments were conducted [12]. First a particular general-purpose dictionary without consideration of the contextual valence shifter terms were under taken. Then, by using two opinions lexica: the general-purpose and the domain specific lexicon was achieved [13]. The last experiment was conducted by using the two lexica and the contextual valence shifter terms were considered.

Abreham Getachew [14] conducts research constructing opinion-mining model that classify Amharic opinionated text into positive or negative using 616 review texts collected from Ethiopian Broadcasting Corporation, diretube.com as well as habesha.com site. Accordingly, the performance of classification shows, NB is equal to 90.9% accuracy of performce, DT and ME the performance of 83.1% and ME 89.6% respectively.

Mohammed Tune [15] proposed a graph-based opinion summarizing system whose vertices contain message objects or topic under discussion and its reply nodes that are labeled with opinion polarities. The opinions were obtained from social media and annotated by experts. The model extracts the summary of that opinion's polarity from corpus of opinion-oriented graph. The experimental results provided information that cannot be provided by solely straightforward text mining or Computational linguistics. Because, the methodsof computational linguistics cannot notice the semantic relationships between more than two sentences.

3. Methods

3.1. Data Source and Data Collection Methods

The datasets used for conducting experiment was collected from ORTO website (http://www.orto.gov.et), at news domain. A total of 400 review datasets were used to conduct the experiment.

3.2. Design Approach

Rule-based approach was used to detect the polarity of opinion and summarization. This rule can be made based on nature of Afaan Oromoo language, to detect the aspect of the opinions. Then lexicon words database that express opinion words was built. To detect the polarity of the opinions, count the number of positive and negative opinions available in each aspect and displays the summary of opinions in bar chart.

3.3. Implementation Tools

Java-programming language is used for implementing the algorithm to creating user interface to facilitate aspect-based opinion summarization and MySQL to create database to data repository like keyword or lexicon of the opinion for detecting the polarity of opinion.

3.4. Evaluation Techniques

Precision, recall, F-Measure, and user acceptance testing are used to evaluate the performance of the system.

Proposed Architecture

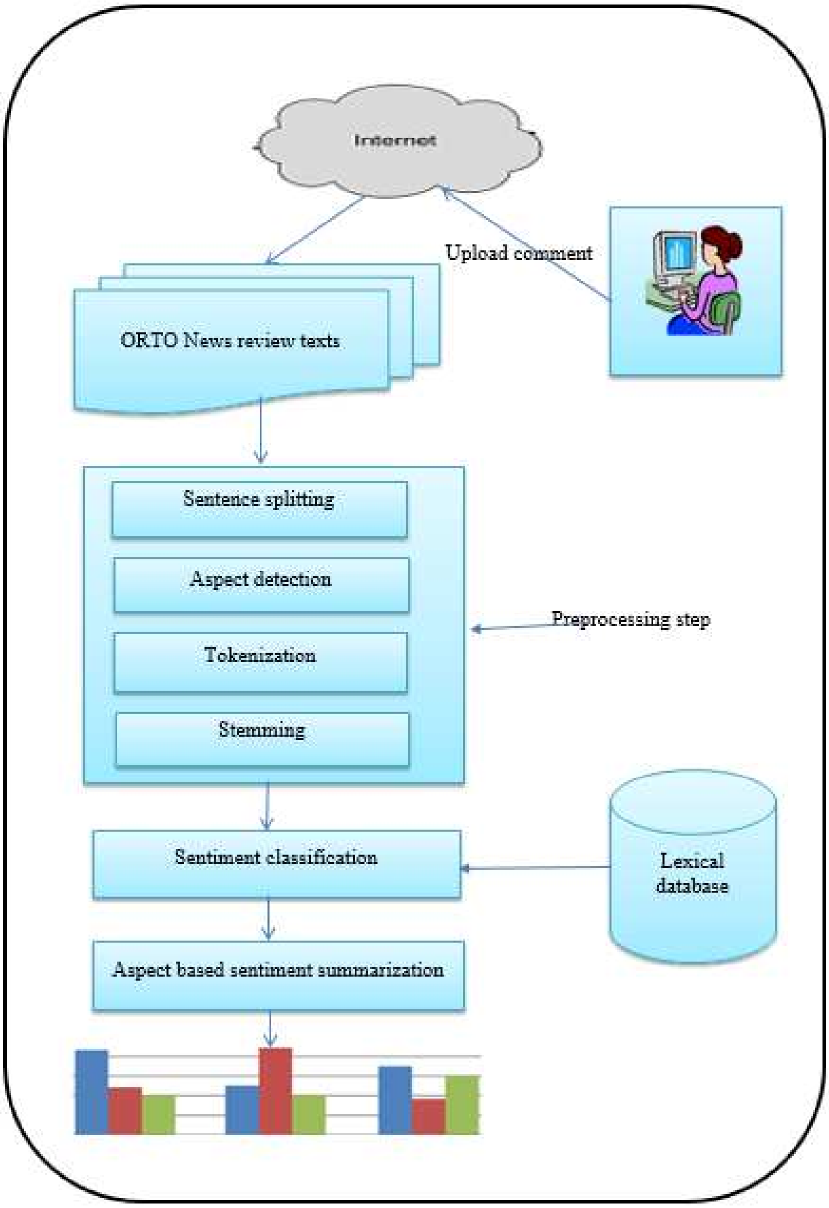

Figure 1. *Architecture of proposed system.*

Given Oromia Radio and Television news review text as an input, the opinionated texts are preprocessed to prepare data set for training and testing. The detail of each step is discussed below.

3.5. Sentence Splitter

The text document is divided into individual sentences using a sentence splitter. This was demonstrated by taking into account punctuation that is used at the end of sentences, such as full stops (.) and exclamation points (!). Sentence splitters, also known as sentence segments', separate text according to specific delimiters into individual sentences. [16].

Sentence Split source code

```
String[] titles = txtDisplay.getText().split("Title: ");
String[] opinion = titles[t].split("\n");
```

3.6. Feature or Aspect Detection

Aspect expressions for a product and/or service review in the sentence that are nouns, and noun phrases are explicit aspect expression. For instance, 'sound' in "The sound of this phone is clear" is explicit aspect expression. The other one is implicit expressions of aspect, as a result of they imply some aspects. For instance, "large" is an implicit aspect expression in "This phone is too large". After identifying aspect as noun and noun phrase then its frequency is found and the most frequent aspects are chosen. The infrequent aspects are unwanted. Algorithm depicted as follows.

Aspect Identification source code

```
for (int i = 0; i < opinion.length; i++) {
```

```
if(i==0){
category[t]=opinion[0]; } }
```

Stemmer

Stemmer is a component which reduce morphological variations of the words into root or base form. In morphologically complex languages such as Afaan Oromoo, a stemming lead to significant improvements in opinion mining. Only the root word and its polarity are saved in the dictionary and retrieved when needed. In this study we adopt Debala [17] stemmer.

3.7. Sentiment Classification

This component indicates the Orientation of a viewpoint on each aspect, which is the polarity score of the viewpoints on the aspect. In this study, opinion words—a dictionary or keyword of opinion words that includes both positive and negative phrases are given an important score using the review corpus. If a word appears in a dictionary or as a keyword and its related value is positive, then the opinion term is positive. In the same way, if a phrase is located in a dictionary and its related value is negative, the opinion term is also negative. For negation, suppose if a word is found in the dictionary and if its' corresponding value is positive and negated one times, then the opinion term is negative. If the term is found in dictionary and if its' corresponding value is negative and negated one times, then this opinion term is positive. Polarity word is the term which can express the opinions towards an object likewise "gaarii" (good) which expresses positive opinion, and "gaarii miti" (not good) or 'gadhee' (bad) which express negative opinion. Afaan Oromo keyword or lexicon database, which contains Afaan Oromo opinion words, the sentiment predictions help in summarizing the general sentiments polarity of the product's aspect/feature.

Once the key word is saved in dictionary along with its polarity, the following line of code select the key word from the dictionary and save on "keyword" and its type or polarity on "type".

To detect the polarity of opinions from huge number of opinions file based on aspect, first user accept the opinions review file then detect the polarity of opinions then initialize number of positive and negative polarity of opinions. Then Get keyword from lexical database to cross check with the words available in the reviews then check negation after and before the keyword then determine the polarity of opinions, if polarity is positive, count the positive polarity of opinions which here under each aspects and also if the polarity is negative, count the negative polarity which here under each aspects and finally summarize the total polarity available in the opinions based on aspects available by using bar chart for visualization.

3.8. Lexical Database

Lexical database, which have root of opinion words like gaarii (good), bad (gadhe) etc that is used for polarity of sentiment classification in the review news texts.

3.9. Aspect-Based Opinion Summarization for Visualization

Most sentiment analysis programs aim to analyze the opinions of a sizable number of users. A decision on a specific problem cannot be effectively made based on a single person's perspective. This indicates that a summary of opinions in some way is required. Aspect-based or feature-based opinion summaries are ways of summarizing information that is based on the features of a good or service that is being offered. Positive and negative score of aspects are separately aggregated; as a result, a combined positive and negative score of aspects is obtained. An algorithm for summarizing opinions based on aspects takes the polarity of positive and negative opinions as input, counts the total polarity of positive and negative opinions, and then summarizes the polarity of opinions based on aspects. The aggregated positive score and negative score of user-generated content has been used for generating summary using visualization tools, which is easy to understand user's opinion in structured form. In this study, bar chart is drawn for generating summary of user's news review about ORTO.

3.10. Building Sentiment Lexicon

To develop sentiment lexicon we use different sources like Amharic sentiment lexicon developed by Selama Gebre meskel [11] and English word net, which contains positive and negative opinion words by translating the word into Afaan Oromo language based on Afaan Oromo language rules. Additionally, we use hard copy of Afaan oromo dictionary "Galmee Jechoota Afaan Oromo" made by Hinsene Mekuria to collect additional Afaan oromo sentiment terms. Totally 1027 Afaan oromo opinion terms where 521 of termss are positive (+) and the rest of 506 terms are negative (-) terms were collected. Lastly, these Afaan oromoo opinion terms were validated by professional students from the Linguistic Department of Addis Ababa university.

Designed user opinion summarization

The designed user opinion mining and summarization works as follow. First user run the developed prototype and then the interface is displayed then users can start interacting with the system, user can feed the document in to the system using browser FileChooser options and then import the selected document into text display field by using import the browsed file button the press the summarize button then aspect summarization of bar chart based opinion is displayed that can be easy for visualization by user and organization to get concise summary of opinion for making decision. Opinion mining and aspect-based summarization techniques are evaluated on 400 (70 Oduu ispoortii/sport news, 56 Oduu oolmaa gabaa /business, 48 Oduu haala qilleensa/metrology, 18 Oduu siyaasa/political, 60 Oduu saayiinsii fi technoloojii/science and technology, 55 Oduu fayyaa fi barnoota/health and education, 47 Oduu aadaa fi turizimii culture and tourism and 46 Oduu biyyaaleessa/international news aspects).

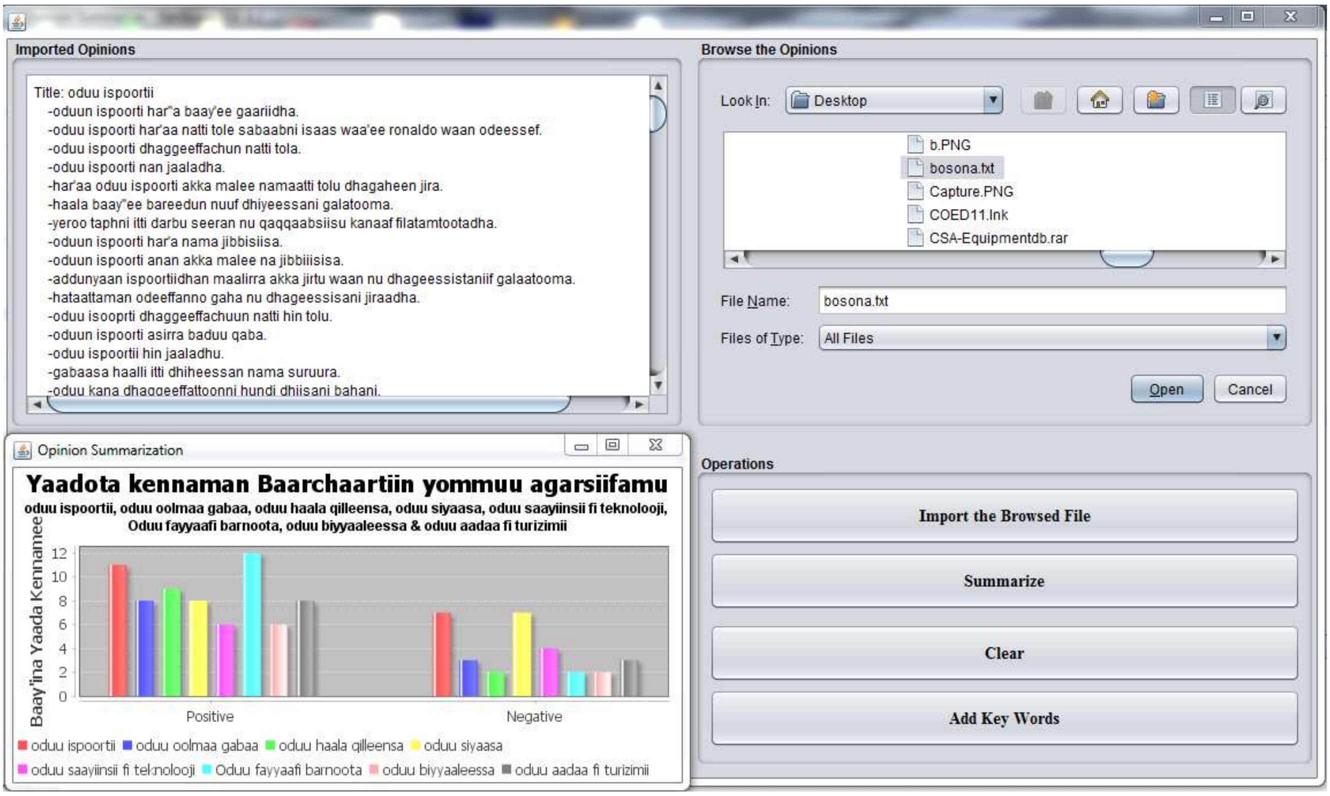

Figure 2. *Feature level opinion mining and summarization with bar chart.*

System performance testing

Three employees from ORTO those who are working as customer opinion officer and opinion analyst, for purpose of evaluating or testing accuracy of developed prototype were selected. Four hundred opinion reviews were given for each evaluator. The result is printed below in table 1. difference between human and system opinion classification.

Table 1. *Experimental result for opinion mining and aspect-based opinion Summarization.*

	Class	Precision	Recall	F-measure
Automatic opinion mining and aspect level opinion summarization	Positive	0.90	0.871	0.885
	Negative	0.87	0.897	0.88

The other experiment done to measure the performance of summarization process by comparing manpower or manually classifying opinions and aspect opinion summarization against developed system prototype. Three persons for manual classification and aspect-based opinion summarization were selected. The result is shown in the following table 2.

Table 2. *Comparisons between manpower and automatic opinion mining and aspect-based opinion summarization.*

Classifying and summarizing aspect level opinions	Time	Man power	Review document
Manpower classification and summarization	60 mins	4	400 opinions
Automatic classification and summarization	2 mins	1	400 opinions

User acceptance testing

Ten (10) people are chosen for the user acceptance testing process, out of which five (5) are domain experts from the ORTO staff and five (5) are other users who speak the Afaan Oromoo language. These individuals will evaluate the prototype and be given explanations on it in order to prevent user awareness gaps about the prototype. After evaluators (i.e. users who have been chosen to evaluate the system engage with the proposed system through test cases that contain parameters and regulations that are comparable to those in the proposed prototype. They then provide feedback via questionnaires that have been developed. Answers to each of the seven open-ended questions are listed in the table. 3.

Table 3. *User acceptance testing.*

No	Questions	Poor	Fair	Good	Very good	Excellent	Average
1	Is the proposed prototype easy to use and interact with?			2	5	3	4.1
2	Is the prototype attractive?			1	5	4	4.3

No	Questions	Poor	Fair	Good	Very good	Excellent	Average
3	Is the prototype efficient in time?				6	4	4.4
4	Is the prototype giving the right opinion polarity classification based on aspects?			1	7	2	4.1
5	What do you think the contribution of the system in the study area?			1	6	3	4.2
		Total Average					4.2

As it is shown on the table 3, average performance of the proposed prototype according to the evaluation result filled by domain experts is 4.2 out of five (5) which means 84%, which is a very good achievement by user acceptance testing.

4. Result and Discussion

The experimental finding demonstrates the usefulness of the suggested model in identifying opinion traits and calculating the polarity of opinion words for the ORTO in news domains. It was able to determine opinion words with 90% precision and 87.1% recall in the positive class, 87% precision in the negative class, and 89.7% recall overall. Using a dataset of 400 gathered reviews and eight attributes, the outcome is produced.

Table 4. *Confusion Matrix of System Performance Testing.*

	System prototype output		
		Positive	**Negative**
Domain expert judgment	Positive	180	20
	Negative	26	174

The accuracy of the proposed system is 88.3% for aspect-based opinion summarization, as shown in the confusion matrix (table 4). 46 of the 400 reviews of viewpoints are incorrectly categorized. Aspect-based opinion summarization and automatic opinion mining have been compared to human labor, and the results demonstrate that the proposed model is more efficient in terms of time and labor costs. In this study, the suggested solution reduces manual clarification time by individuals by more than 96%.

The challenges in this study are users give their opinion in context based or/and indirect manner which is difficult to identify the polarity of opinions. For example: in the review "jalqaba kan ormaa odeessuu irraa osoo waa'ee keenya baree gaarii dha. /before talking about other it's better to know our service. Even though, the opinion expressed is negative, but the proposed prototype labeled it as positive because the reviewer used positive opinion term 'gaarii' (better), to express negative opinions toward the Oromia radio and television organization in news domain. Sometimes words are misspelled which the system cannot understand.

5. Conclusion and Recommendation

Web 2.0 innovations paved the stage for the quick expansion of user-generated content online. A person with an opinion can upload user-generated content online through a variety of blogs, review sites, social media platforms, discussion forums, and websites run by organizations and non-organizations. The vast amount of user-generated content might be challenging to manually gather, comprehend, summarize, and evaluate. Therefore, a decent tool is needed to automatically mine and summarize sentiments at the aspect level in reviews of a certain product or service in order to comprehend the opinion holders' attitudes toward a particular entity or item. In this study, the problem of identifying and extracting aspect/features and determining opinion polarity and then generating the summary is done using rule-based approach. The result of the experiment, shows that performance of 90% precision and 87.1% recall in positive class, 87% precision for negative class, 89.7% recall and is 88.3% system accuracy is achieved in the determination of aspect-based opinion summarization. Bar chart, based visualization is attempted, which is also accepted by user into very good user acceptance ratio. Accordingly, the system registers promising result. However, since the system did not consider context it is easily affected by indirect comment given by the user about ORTO'S service wrongly spelled words were the challenges to the current study. To extend this study in the context-based opinion mining and summarization can be considered.

References

[1] B. Liu, "Sentiment Analysis and Opinion Mining," Morgan & Claypool Publishers, April 22, 2012.

[2] Bo Pang and Lillian Lee, "Opinion mining and sentiment analysis," Foundation and Trends in Information Retrieval, vol. 2, pp. 1-135, 2008.

[3] B. Liu, "Sentiment Analysis," 5th Text Analytics Summit, Boston, pp. 1-2, June 2009.

[4] Lei Zhang and Bing Liu, "Aspect and Entity Extraction for Opinion Mining," Data mining and knowledge discovery for big data, pp. 1-40, 2014.

[5] G. Vinodhini and R M. Chandrasekaran, "Sentiment analysis and Opinion Mining: A survey," International Journal of Advanced Resarch in Computer Science and Software Engineering, vol. 2, no. 6, pp. 1-11, June 2012.

[6] L. Zhang, "Aspect and Entity Extraction from Opinion Documents," 2012.

[7] E. Akba, "Aspect based opinion mining on turkish tweets," Bilkent university, 2012.

[8] Alasmar, Ahmed M., "Feature Based Approach in Arabic Opinion Mining Using Ontology," 2016.

[9] T. T. Hailu, "Opinion Mining from Amharic Blog," Ethiopia, 2013.

[10] E. Alamerew, "Automatic Annotation of Opinionated Amharic Text For Opinion Mining," Ethiopia, 2016.

[11] S. Gebremeskel, "Sentiment Mining Model for Opinionated Amharic Texts," Ethiopia, 2010.

[12] Noah Smith and David Smit, 2005. [Online]. Available: http://www.cs.jhu.edu/~nasmith/erm/ (Accessed date 4/6/2017). [Accessed 4 February 2017].

[13] Mcsherry, Frank and Marc Najork, "Computing information retrieval performance measures efficiently in the presence of tied scores," Advances in information retrieval, pp. 414-421, 2008.

[14] A. Getachew, "Opinion Mining from Amharic Entertainment Texts," Ethiopia, October, 2014.

[15] M. Tune, "Designing A Graph-Based Opinion Mining Model for Opinionated Text in English, Amharic and Afaan Oromo Language," Ethiopia, February 2015.

[16] "Substitutions and sentence splitting," Pandorabots, 2017. [Online]. Available: http://docs.pandorabots.com/tutorials/substitutions-and-sentence-splitting/. [Accessed 5 4 2017].

[17] D. Tesfaye, "Designing a Stemmer for Afan Oromo Text: A hybrid approach".