Received: March 9, 2021. Revised: April 26, 2021.

INASS

International Journal of
Intelligent Engineering & Systems

http://www.inass.org/

Aspect Based Sentiment Analysis for Explicit and Implicit Aspects in Restaurant Review using Grammatical Rules, Hybrid Approach, and SentiCircle

Rachmad Abdullah[1] **Suhariyanto**[2,3] **Riyanarto Sarno**[2]*

[1]*Deparment of Information Technology Management, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia*
[2]*Deparment of Informatics, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia*
[3]*Faculty Of Computer Science University of Dian Nuswantoro, Semarang, Indonesia*
* Corresponding author's Email: riyanarto@if.its.ac.id

Abstract: TripAdvisor is one of the most popular e-commerce platforms in the tourism sector in Indonesia. TripAdvisor give Traveler Choice Award every year in Indonesia through user reviews. However, online text-based reviews are often associated only with evaluation scores that do not pay attention to the context and meaningful content of the review itself, either explicitly or implicitly. Moreover, the sentence structure of the review can have an impact on the goal of the target sentiment which is nothing but an aspect of the review itself. This research discusses aspect based sentiment analysis for explicit and implicit aspects. This research starting with taking the TripAdvisor website restaurant product review dataset to measure customer satisfaction based on four aspect categories of Ambience, Food, Service, and Price. Furthermore, the aspect word extraction and opinion word extraction processes in the case of explicit sentences for simple, compound, complex, and compound-complex sentence structures are carried out using grammatical rule extraction. This research also works on implicit sentence cases for simple sentence structures. Aspect categorization process uses hybrid approach. Aspect and opinion keyword extraction process uses the ELMo-Wikipedia. Then WordNet and TF-ICF are used to expand the meaning of aspect and opinion that has been taken. The last stage is the aspect based sentiment analysis process, both explicit and implicit sentences using SentiCircle. This research can produce two evaluations of sentiment classification, namely positive and negative. The results of the aspect extraction obtained an evaluation of the aspect categorization for each precision 0.82, recall 0.87, and f1-measure 0.86. Meanwhile, the results of the sentiment analysis showed that the respective evaluations for precision 0.87, recall 0.92, and f1-measure 0.89.

Keywords: E-commerce, Aspect based sentiment analysis, Grammatical rule extraction, Aspect categorization, Hybrid approach, ELMo-wikipedia, WordNet, TF-ICF, SentiCircle.

1. Introduction

Aspect Based Sentiment Analysis (ABSA) is an approach to find out a user's opinion from what aspects he likes and dislikes [1]. Here, nouns and noun phrases were often extracted as candidate aspects. It makes opinion words used to extract aspects are rare. ABSA research case development [2] divides ABSA process into three parts, there are extraction and aggregation of aspect, and sentiment analysis. Aspects were extracted based on a manually created set of dependency relationship rules [3]. However, the fact of the extraction does not cover all the possible rules.

In other researches, ABSA is widely applied in cases of explicit sentences, both aspects and sentiments. In the hotel domain, conducted ABSA research using topic modeling and semantic similarity [4, 5]. These researches can solve the sentiment analysis case based on aspects in simple sentence patterns. These researches can generate positive and negative sentiment values too. However, these researches are not work in other structure of sentences. The other research was conducted by taking explicit aspects using rules based on constituency parse by paying attention to word phrases [6]. However, these rules work by assuming

that the aspect word are close to opinion word in a sentence, so if there are multi-aspects and multi-opinions, the rules cannot work properly. Likewise, similar research [7]. All of the above research, focus on the extraction of explicit aspects and opinions without paying attention the structure of sentence.

Recent research for the extraction of implicit aspects has been carried out [8]. This research extracted Indonesian sentences on a pattern of equivalent compound sentences and multilevel compound sentences to obtain explicit and implicit opinions with positive and negative sentiment values. However, this research not discussed implicit aspect based on the stucture of sentence. Therefore, aspect-based sentiment analysis for English is needed to improve previous research methods.

By surveying published works on the application of such sentiment analysis, this research was conducted with a focus on the review of English because most of the people use English at Online Social Network, the scarcity of resources in it, and the challenges associated with research in it. The review is taken from the TripAdvisor website because TripAdvisor is one of the most popular e-commerce platforms in the tourism sector in Indonesia. TripAdvisor give 'Traveler Choice Award' every year in Indonesia with the best hotel, destination, beach, flight, attraction and restaurant categories. TripAdvisor is also equipped with a feature that contains reviews from users. This research uses a dataset from TripAdvisor website restaurant product reviews to measure customer satisfaction based on the Ambience, Food, Service, Price aspect categories.

This research assumes that the structure of a sentence can affect the process of extracting aspect and opinion words for sentiment analysis. For example: how to extract a review consisting of only 1 opinion word "Overpriced!" for sentiment analysis. Another example: how to extract a review from a compound-complex sentence for sentiment analysis, "I bought my food about an hour ago and I have to say that I am very satisfied", where the pair of aspect word “food” and opinion word “satisfied” are not close together in the sentence. The previous researchs mentioned above have not discussed these issues.

This research provides suggestions on how to build a rule for the aspect categorization process using a hybrid approach. Hybrid approach concept is using grammatical rule extraction, Elmo-Wikipedia, WordNet, and Term Frequency-Inverse Component Frequency (TF-ICF) to handle the extraction of aspect and opinion words in cases of explicit and implicit sentences. Hybrid Approach is used to identify more accurately the pair of aspect and opinion words in four types of sentence structure, namely: simple, compound, complex, and compound-complex sentences. The focus of this research is taking aspect and opinion words from noun phrases and adjective phrases in explicit and implicit aspects sentences. Then, how to extract aspect and opinion words in simple, compound, complex, compound-complex sentence structures.

The research starts with extraction of explicit sentence case in simple sentence that has explicit aspect and explicit opinion which labeled as Si-AEOE. Furthermore, this research works on the case of compound sentence that has explicit aspect and explicit opinion which labeled as Co-AEOE, complex sentence that has explicit aspect and explicit opinion which labeled as Ce-AEOE, and compound-complex sentence that has explicit aspect and explicit opinion which labeled as CoCe-AEOE. The next case carried out in this research is the extraction of implicit sentence case in simple sentence that has implicit aspect and explicit opinion which labeled as Si-AIOE.

In this research, to taking the aspect and opinion word in the case of explicit and implicit sentences, a combination of lexicons is used, namely Opinion Lexicon [1] and Implicit Lexicon [9]. In the assessment of the sentiment analysis evaluation, this research uses Senticircle to extract opinions according to the context of the aspects with positive and negative sentiment values.

2. Related theory

This research explains several theories in this section as follows.

2.1 Aspect variables of restaurant

In previous research about restaurant review [10] which discussed about expectation from customer, thus another research about perception of service quality to customer and also research [11] which discussed about implicit aspect detection. This research measures the restaurant satisfaction value based on four aspect categories reviewed, namely: 1)Ambience, 2)Food, 3)Service, and 4)Price. The variables of these aspect categories are shown in Fig. 1.

2.2 Keyword for restaurant aspect category

Four aspect categories in this research are used as the basis for restaurant ratings from buyers to leave reviews through the TripAdvisor application. We also use aspect keywords from Wikipedia to take aspect and opinion word from these reviews. We use several wikipedia pages to extract keywords for restaurant aspect category.

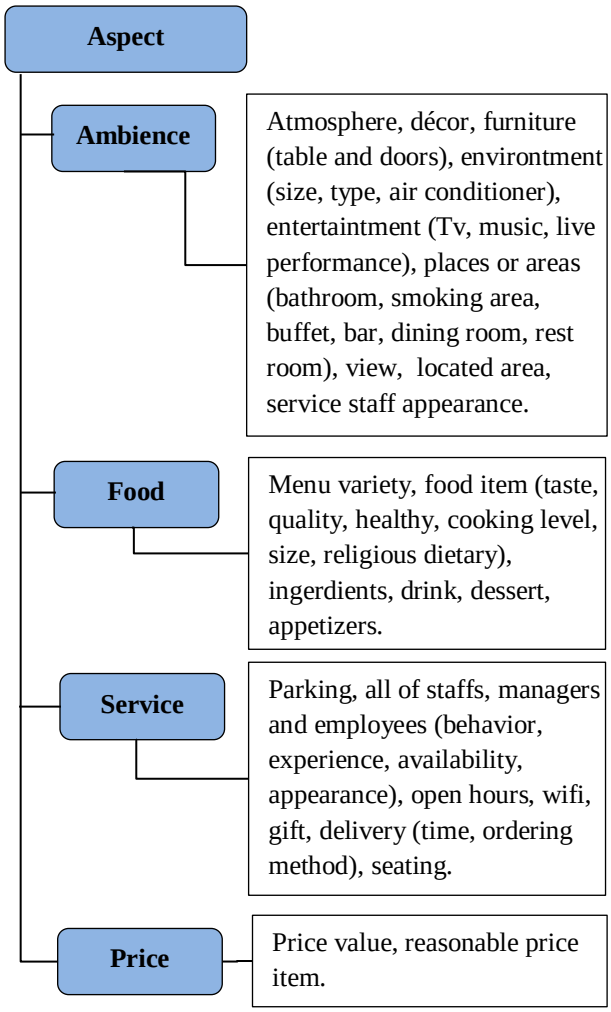


Figure. 1 Categories and variables of aspect

2.3 Pre-processing

In general of natural language processing, these are steps for pre-processing as shown in Fig. 2.

2.4 Grammatical rule

In recent natural language processing, several related researches about grammatical rule extraction have been carried out to extract aspect and opinion words in review. Based on an explanation of word similarity in previous researched, a word has 2 kinds of similarity [12]**.** Grammatical rule extraction works based on similar character sequences to takes lexically similar words and based on the same meaning to takes semantically similar words. In this research, Grammatical Rule is used for sentence extraction to identify and split the sentences become Si-AEOE, Co-AEOE, Ce-AEOE, CoCe-AEOE, and Si-AIOE. The grammatical rule uses opinion lexicon and implicit opinion lexicon for taking the aspect word and opinion word which identified as Noun Phrase and Adjective Phrase.

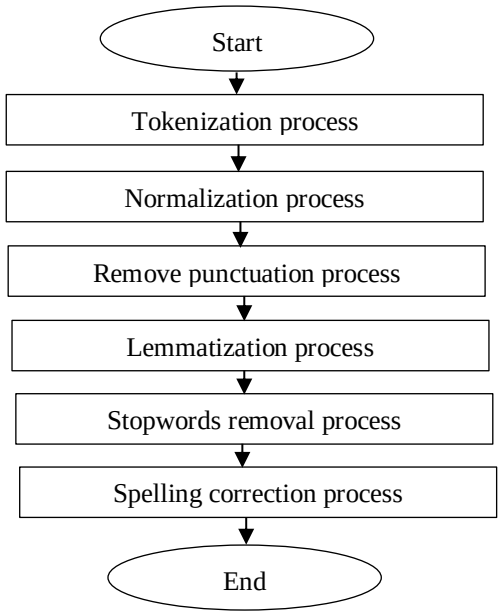


Figure. 2 Pre-processing process

2.5 Term Frequency-Inverse Cluster Frequency (TF-ICF)

Term Frequency-Inverse Cluster Frequency (TF-ICF) [13] is used to calculate the term weight of a word against document of each cluster. For brevity, the ICF value of word i (ICF_i) is calculated based on cluster numbers that containing word i (cfi) on exist cluster numbers (c). The ICF equation is shown in Eq. (1). Then TF-ICF weighting of word i $(TF - ICF_i)$ is shown in Eq. (2). Eq. (2) is the product of word i frequence in cluster j $(TFji)$ and the ICF value of word i (ICF_i).

$$ICF_i = 1 + \log \frac{c}{cf_i} \quad (1)$$

$$TF\text{ - }ICF_i = TF_{ji}\ x\ ICF_i \quad (2)$$

2.6 Elmo

Elmo [14] is used in this research to represent context of word in aspect keywords. We use Elmo because it can make a word be able represent many vector depends on sentence context. Input for Elmo in this research is word and sentence. So, we used TensorFlow tools [15] to convert a word into 1024-dimensional vector and convert a sentence into array of 1024-dimensional vector in Elmo implementation.

2.7 Wikipedia

In general, Wikipedia [16] is used to represents text meaning using machine learning for sentiment analysis in explicit case. Therefore, this research proposes several Wikipedia pages as keywords of

aspect categories that have explained in Sect. 2.2.

2.8 WordNet

WordNet [17] is used to obtain synonyms and antonyms of aspect and opinion words for aspect categorization process [18].

2.9 Semantics

Semantics is the research about the meaning of words. The meaning of the word in a sentence could be in form of denotation, sense, lexical, structure, categorematic, or syncategorematic [19]. Therefore, semantic similarity is needed to calculates distance of words on the meaning of word [20].

Thus, the angle from two word vectors which represent similarity distance the first word (wi) until the second word (wj) will be measured using cosine similarity [21]. The result of cosine similarity ranges is -1 until 1 as shown its equation in Eq. (3). $\sum_{m=1}^{K}$ is the iteration numbers from m to K word.

$$Similarity(w_i, w_j) = \frac{\sum_{m=1}^{K} {w_i}^m {w_j}^m}{\sqrt{\sum_{m=1}^{K} ({w_i}^m)^2}\sqrt{\sum_{m=1}^{K} ({w_j}^m)^2}} \quad (3)$$

2.10 Opinion Lexicon

Opinion Lexicon [22] is used in this research because its contain two opinion lists, namely positive and negative opinion words, around 6800 words.

2.11 Implicit Aspect Lexicon

Implicit aspect lexicon [9] is used in this research because its contain nine implicit aspect categories which represent the aspect categories variables as explained in Sect. 2.1.

2.12 SentiCircle

SentiCircle [23] is used to capture semantics and sentiments based on contextual of words in a sentence. In this research, we use SentiCircle because it is in line with the distributional hypothesis about contexts and meanings of words in a sentence [24]. SentiCircle model, as shown in Figure 3, is a model that contains a coordinate of polar. The model is assumed to the processes below.

1. Generate correlation degree of word term ($TDOC$)
2. Calculate prior sentiment score
3. Calculatex_i and y_i point.

SentiMedian is method for extract sentiment from SentiCircle given point. The sentiment polarity of the context term which calculated by SentiMedian could be positive or negative as shown in Eq. (8).

$$TDOC(n, c_i) = f(c_i, n) \times \log \frac{F}{F_{c_i}} \quad (4)$$

$$\theta_i = Prior_Sentiment(c_i) \times \pi \quad (5)$$

$$x_i = r_i Cos\theta_i \quad (6)$$

$$y_i = r_i Sin\theta_i \quad (7)$$

$$g = min_{g \in R^2} \sum_{i=1}^{j} \|p_i - g\|^2 \quad (8)$$

where,

d = document
n = word term of opinion
c_i = word context of opinion
F = frequency of n in d
F_{c_i} = frequency of c_i in d
$f(c_i, n)$ = frequency of joint occurrences between n and c_i in d
$Prior_Sentiment$ = polarity value
θ_i = c_i degree (in radian)
x_i = c_i position in x axis
y_i = c_i position in y axis
p_i = c_i position
g = Position of n (x_n and y_n)

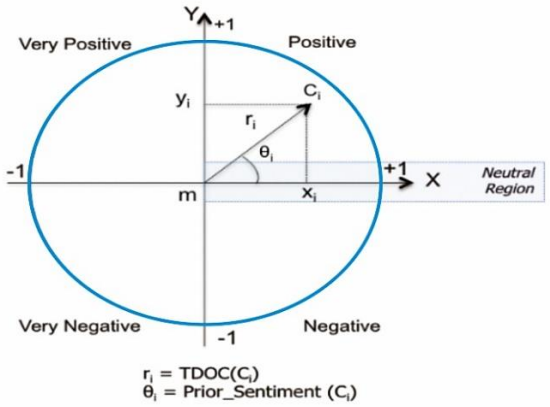


Figure. 1 SentiCircle mode

		Actual Values	
		Positive	Negative
Predicted Values	Positive	True Positive (TP)	False Negative (FN)
	Negative	False Positive (FP)	True Negative (TN)

Figure. 2 Confusion matrix

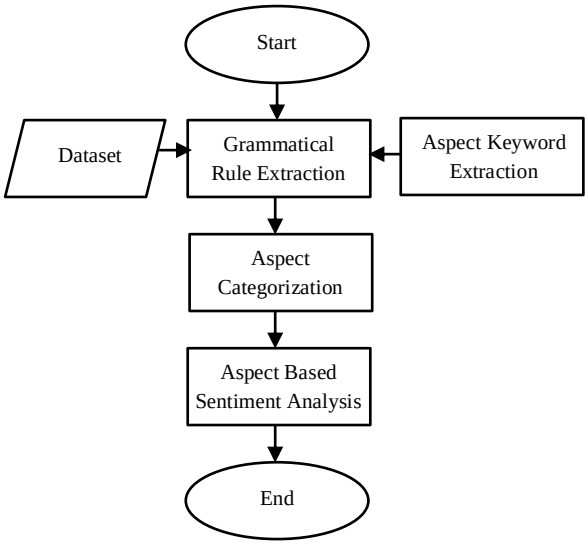


Figure. 3 System flow diagram

2.13 Evaluation

Testing aspect categorization and sentiment analysis uses three evaluation methods, namely Precison, Recall, and F1-measure. To calculate this evaluation, the first step is to form a confusion matrix [20] as in the Fig. 4.

$$Precision = \frac{TP}{TP+FP} \quad (9)$$

$$Recall = \frac{TN}{TP+FN} \quad (10)$$

$$F1 - measure = \frac{2*P*R}{P+R} \quad (11)$$

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \quad (12)$$

3. Research method

This research starts with pre-processing the dataset to make relevant data information. The relevant data is became input in next module, namely Grammatical Rule Extraction (GRE). In GRE module, the relevant data is processed into several structures of sentence. Keywords from Wikipedia are extracted as the input in GRE module before it will be processed in aspect categorization (AC) module. Aspect keywords are resulted from Wikipedia links in Figure 6. In AC module, the aspect and opinion terms are extracted using hybrid approach. Hybrid approach is combined using grammatical rule extraction, WordNet, TF-ICF, and Semantic similarity. Then, aspect and opinion words are extracted based on aspect categories. The next module is Aspect-based Sentiment Analysis (ABSA). In this module, the dataset based on aspect categories is prepared as input for ABSA module using SentiCircle method. Each accuration, precision, recall, and F1-measure is used to measure and evaluate for AC module and ABSA module in this research.

Table 1. Dataset representation

ID Review	ID Sentence	Review
1004293	0	Judging from previous posts this used to be a good place, but not any longer.
1004293	1	We, there were four of us, arrived at noon - the place was empty - and the staff acted like we were imposing on them and they were very rude.
1004293	2	They never brought us complimentary noodles, ignored repeated requests for sugar, and threw our dishes on the table.
…	…	…

3.1 Dataset

Dataset in this research is obtained from the public dataset from Semeval 2015 [25]. This dataset has been annotated, both the category aspect annotation and the sentiment polarity annotation. The dataset representation is shown in Table 1.

3.2 Aspect keyword extraction

3.2.1. Pre-processing

Pre-processing stage in this research is done with the results as follow: 1) convert text to lower case, 2) removes the set of symbols, 3) removes the stop words, 4) reduce inflectional forms, and 5) limit of minimum word.

3.2.2. Wikipedia

This research proposes several Wikipedia page links that correspond to aspect categories that shown in Fig. 6.

3.3 Grammatical Rule Extraction (GRE) module

The process of Grammatical Rule Extraction (GRE) module in this research was adopted from Poria rule-based [26], which was further developed based on Noun Phrase and Adjective Phrase according to several structures of sentence that mentioned in Sect. 1, namely Si-AEOE, Co-AEOE, Ce-AEOE, CoCe-AEOE, Si-AIOE. The detailed steps of GRE module use the basic dependencies parser and the Enhanced ++ dependencies parser from Stanford CoreNLP. This research uses two general rules to parsing the sentence. It is as follows:

- Rules for sentences having a Noun Phrase (NP).
- Rules for sentences do not have a Noun Phrase (NP) and have an Adjective Phrase (AP).

3.4 Aspect and opinion word expanding

3.4.1 Expanding using WordNet

Aspect and opinion words which extracted in GRE module is done if they have complied with aspect keyword and lexicon. Here, WordNet is used to expand them if they are not complied with aspect keyword and lexicon. Flow diagram of expanding process using WordNet can be seen in Fig. 7.

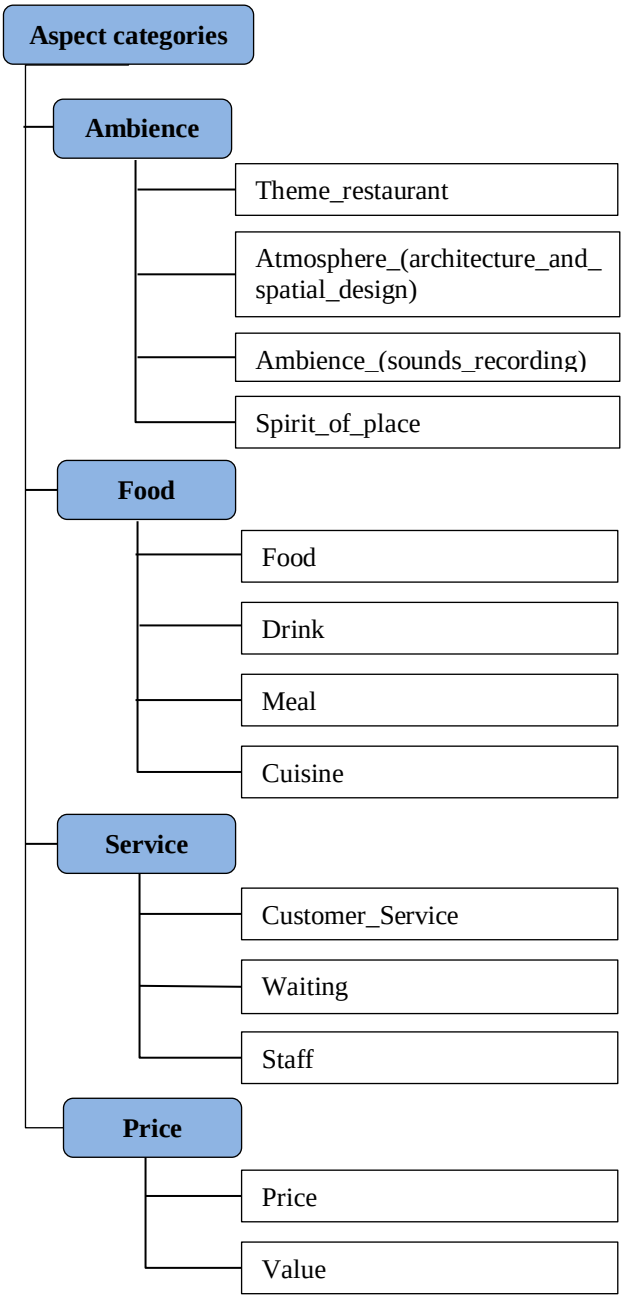


Figure. 4 Wikipedia page titles

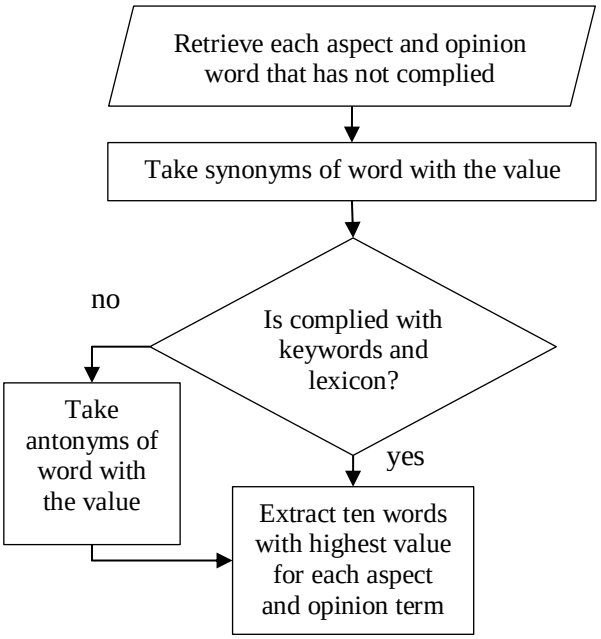


Figure. 5 Expanding process using WordNet

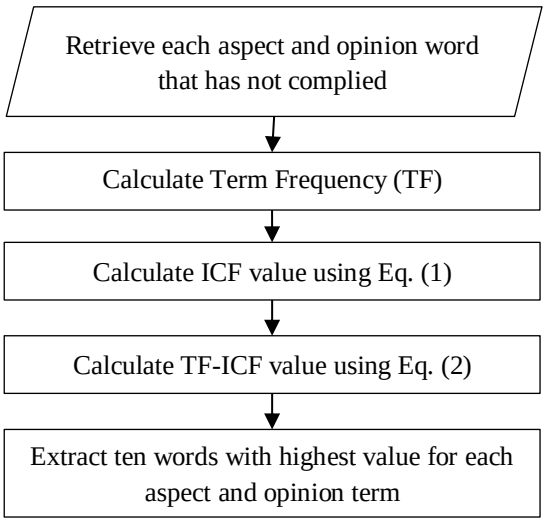


Figure. 6 Expanding process using TF-ICF

3.4.2 Expanding using TF-ICF

The next aspect and opinion word expanding is done using TF-ICF. As same as WordNet, TF-ICF is used to expand aspect and opinion words too. TF-ICF in this research expand the meaning of aspect and opinion words extracted with document in the review that has been pre-processing. TF-ICF formula is described in Eqs. (1) and (2). Flow diagram of expanding process using TF-ICF can be seen in Fig. 8.

3.5 Aspect Categorization (AC)

Aspect categorization (AC) between keywords and terms of aspect and opinion words is calculated using semantic similarity. The values of the aspect and opinion keywords are determined using Elmo-

Wikipedia. Thus values of the aspect and opinion words are determined from GRE module with WordNet and TF-ICF expansion.

The AC process is divided into 3 stages, namely AC1, AC2, and AC3. AC1 method using Grammatical Rule, ELMo-Wikipedia, and Semantic. AC2 method using Grammatical Rule, ELMo-Wikipedia, TF-ICF 20%, and Semantic. AC3 method using Grammatical Rule, WordNet, ELMo-Wikipedia, TF-ICF 100%, and Semantic.

3.6 Aspect Based Sentiment Analysis (ABSA)

Aspect Based Sentiment Analysis (ABSA) conduct several experiments to get the best ABSA performance. First, SentiCircle method is used on ABSA1. Then, ABSA2 using SentiCircle and Opinion Lexicon. Finally, ABSA3 using SentiCircle, Opinion Lexicn, Implicit Aspect Lexicon, and Co-occurrence of context terms.

4. Result and analysis

4.1 Result of GRE module

This research makes classification of review using GRE module before take the aspect and opinion word in aspect categorization process. The evaluation score result of GRE module in every sentence structure shown in Table 2. Table 2 shows that GRE module can extract well for aspect and opinion words in 5 types of sentence structure.

4.2 Result of AC

Table 3 shows this research result about comparison of AC performances from AC1, AC2, and AC3. WordNet and TF-ICF 100% can expand the aspect and opinion words in sentence to be paired with the existing Wikipedia keywords. They can predict aspect and opinion words more accurately. For more detail, the evaluation score of AC performance shown in Table 4. The best AC approach is AC3 with F1-measure value 0.86.

It can be seen from Table 4 that Wikipedia alone or with only adding TF-ICF still not be able used as

Table 2. Result of GRE module performance

Structure	Accuracy Score of Extraction	
	Aspect word	Opinion word
Si-AEOE	0.99	0.99
Co-AEOE	0.98	0.89
Ce-AEOE	0.97	0.86
CoCe-AEOE	0.89	0.81
Si-AIOE	0.87	0.90

Input: Sentence 'S'
Output: Set of Si-AEOE, Co-AEOE, Ce-AEOE, CoCe-AEOE, Si-AIOE

Start

1. Create Basic and Enhanced++ dependencies of S using Stanford CoreNLP;
2. Generate Modified dependencies of S by taking subject clauses, 'nsubj' or 'nsubjpass' from Enhanced++ dependencies and other linked parts from Basic dependencies;
3. Generate dependencies marked as 'acl', 'appos', 'advcl', 'cc', 'ccomp', 'dep', 'mark', 'parataxis', 'cop', 'conj', and 'ref';
4. When the sentence does not have 'conj' and 'mark' dependencies then:

a. If an active token y has a subject noun relationship with a word x then generate S as Si-AEOE and then:
   i. Running the Noun Phrase (NP) rules
   ii. Running the Adjective Phrase (AP) rules

b. If an active token y has not a subject noun relationship with a word x and x exist in implicit aspect lexicon then generate S as Si-AIOE and then x is extracted as an aspect.

5. When the sentence has 'conj' dependencies and does not have 'mark' dependencies then generate S as Co-AEOE and then:
   i. Running the Noun Phrase (NP) rules
   ii. Running the Adjective Phrase (AP) rules
   iii. If an aspect term x has co-ordination or conjunct relation with token y, then y is extracted as an aspect too.
6. When the sentence does not have conj dependencies and have mark dependencies then generate S as Ce-AEOE then:
   i. Running the Noun Phrase (NP) rules
   ii. Running the Adjective Phrase (AP) rules
   iii. If y is extracted as an aspect and y has noun compound modifier x, then the aspect x-y is extracted and y is removed from the aspect list.
7. When the sentence has conj and mark dependencies then generate S as CoCe-AEOE then:
   i. Running the Noun Phrase (NP) rules
   ii. Running the Adjective Phrase (AP) rules
   iii. If a token y_0 has direct object relation with a token y, if y is extracted as an aspect or y has noun compound modifier x, then if y_0 has co-ordination or conjunct relation with another token y_1, then if y_1 is in open clausal complement relation with a token y_2, then if y_2 is in open clausal complement relation with a token y_3 whose POS is Adjective, then the aspect y-x is extracted if y-x exists in the opinion lexicon.
8. Rearrange the words found in step 4 until step 8 to generate the complete sentence.

End

Figure. 7 AC process rule algorithm

Table 3. Result of AC performance

AC Performance		
AC Approach	**Method**	**F1-measure**
AC1	GRE + ELMo-Wikipedia + Semantic	0.80
AC2	GRE + WordNet + ELMo-Wikipedia + TF-ICF 20% + Semantic	0.84
AC3	**GRE + WordNet + ELMo-Wikipedia + TF-ICF 100% + Semantic**	**0.86**

Table 4. Comparison of aspect categorization process

	Review	**Aspect term extracted**	**Aspect categorization results**
1	Excellent food for great prices	food, prices	AC1 : {food:excellent} {prices:great} AC2 : {food:excellent} {prices:great} AC3 : {food:excellent} {prices:great}
2	the pizza was overpriced and soggy.	pizza, pizza	AC1 : {pizza:overpriced} AC2 : {pizza:overpriced} AC3 : {pizza:overpriced} {pizza:soggy}
3	The rest of the dim sum, though pricey by Chinatown standards, is worth it.	dim sum	AC1 : {dim sum:worth} AC2 : {dim sum:worth} AC3 : {dim sum:worth}
4	I bought my food about an hour ago and I have to say that I am very satisfied.	food	AC1 : {food:satisfied} AC2 : {food:satisfied} AC3 : {food:satisfied}
5	Overpriced!	overpriced	AC1 : {overpriced:overpriced} AC2 : {overpriced:overpriced} AC3 : {overpriced:overpriced}

a good keyword parameter to taking aspect and opinion words, especially in the case of implicit sentences.

Review “Excellent food for great prices” is a review labeled Si-AEOE so that it can still be extracted properly using Wikipedia keywords. Here, the word "prices" is not exactly the same as the word "price" but semantic similarity still be able take it as an aspect word. The result can be taken from the aspect and opinion pairs of "food: excellent" and "prices: great".

Review “the pizza was overpriced and soggy.” Which labeled as Co-AEOE displayed different results when only using keywords from Wikipedia compared to adding wordnet and TF-ICF extension. It can be seen that AC1 using only the Wikipedia keyword only can generate the aspect and opinion words pair “pizza: overpriced”. This is because the word "soggy" is not found in the lexicon that used so it cannot be taken. cannot take the opinion word "soggy". AC2, In the same case with AC1, if only the TF-ICF word extension is used, the aspect and opinion words pair can be produced as same as with AC1. The reason is the word "soggy" contained in the document, amounting to 20% of the total sentences containing the word "soggy" cannot be taken from the word similarity in the lexicon that used. Meanwhile, these problems can be solved using the AC3 method. This is because the expansion of the word "soggy" using wordnet produces several words that can be taken from the similarity of the word "soggy" with the lexicons used, namely "sloppy" and "muddy". So it can produce a pair of words "pizza: overpriced" and "pizza: soggy".

Review “The rest of the dim sum, though pricey by Chinatown standards, is worth it.” which labeled as Ce-AEOE also shows that Wikipedia can work properly to take the aspect word "dim sum" which when compared to using wordnet and TF-ICF alone cannot take it as in previous researches. This is because "dim sum" is a series of word phrases which many previous researches have not been able to do the extraction case but Poria et al can do it well. So that the word pair "dim sum: worth" can be generated.

Review “I bought my food about an hour ago and I have to say that I am very satisfied.” which labeled as CoCe-AEOE shows that Poria et al have not worked on the case of similar sentences properly due to the limited algorithm. In this research, we generate the novelty of the adoption of the Poria method because we can also extract compound-complex sentence cases from the rule algorithm that we add. So it can produce pairs of words "food: satisfied".

Review “Overpriced!” which labeled as Si-AIOE also shows that the novelty of grammatical rule extraction in this research can works well because it can takes the word "Overpriced" as an aspect. This is an update of the research we are working on. We can

still take aspect and opinion words even though the sentence is a simple sentence where there is only a single word as an independent subject without even a predicate or object. We can take the word overpriced because it is a word that falls under the implicit aspect lexicon. So that it can produce an aspect pair of words and the opinion is "overpriced: overpriced".

4.3 Result of ABSA

The evaluation result of ABSA shown in Table 5. Table 5 shows a comparison of ABSA performance using SentiCircle, namely ABSA1, ABSA2, and ABSA3.

SentiCircle is used to predicts the context of the word. For instance, these are results of 5 review from Table 9 to distinguish the sentiment of a word and its polarity. Fig. (10), (11), and (12) show the example of extraction results from 5 dataset review using SentiCircle to takes aspect and opinion word based on ambience, food, service, and price from the aspect categories. These figures show that SentiCircle can predicts the context of the word in explicit case very well but not for implicit case.

Review 1 result, SentiCircle in each performances ABSA1, ABSA2, and ABSA3 extracts a pair of aspect words and opinions "food:excellent" and "prices:great" properly. Review 2 result is shown in Fig (10), (11), and (12). SentiCircle in ABSA1 extracts the aspect and opinion word pairs "pizza:overpriced" into the FOOD category because the aspect word taken is "pizza" with the sentiment of the opinion word "overpriced" is positive. Furthermore, SentiCircle in ABSA1 has not be able to extracts the aspect and opinion pairing of "pizza: soggy" because it cannot take the word "soggy" as an opinion word. SentiCircle in ABSA2 is similar in implementation to ABSA1, extracts the aspect and opinion word pair "pizza: overpriced" into the FOOD category because the aspect word taken is "pizza" but the sentiment of the word "overpriced" is negative because it uses opinion lexicon. SentiCircle in ABSA2 also has not be able to extracts the aspect and opinion pairing of "pizza: soggy" because it cannot take the word "soggy" as an opinion word. SentiCircle in ABSA3 shows more accurate result compared with ABSA1 and ABSA2. This is because ABSA3 uses the additional implicit aspect lexicon method so it can extract the aspect and opinion pairs of "pizza: overpriced" into the PRICE category because the opinion word "overpriced" is included in the implicit lexicon which is marked with the PRICE aspect category with the sentiment of the word "overpriced" has negative value because of using opinion lexicon. Furthermore, SentiCircle in ABSA3 also can extract aspect and opinion word pair "pizza: soggy" into the FOOD category from the aspect word "pizza" and negative sentiments from the opinion word "soggy". This is because ABSA3 uses a co-occurrence of context word which can take the similarity of the word "soggy" with the opinion words "sloppy" and "muddy" in opinion lexicon.

Table 5. Result of ABSA

ABSA Performance		
ABSA Approach	**Method**	**F1-Measure**
ABSA1	SentiCircle	0.84
ABSA2	SentiCircle + Opinion Lexicon	0.86
ABSA3	**SentiCircle + Opinion Lexicon + Implicit Aspect Lexicon + Cooccurrence of context terms**	**0.89**

Review 3 result shows that SentiCircle on ABSA1 and ABSA2 cannot extract the aspect and opinion word pair "dim sum:worth" because it cannot recognize the aspect word phrase "dim sum". However, in ABSA3 this problem can be solved with the result of aspect word "dim sum" as the FOOD aspect category and the positive sentiment value of the opinion word "worth".

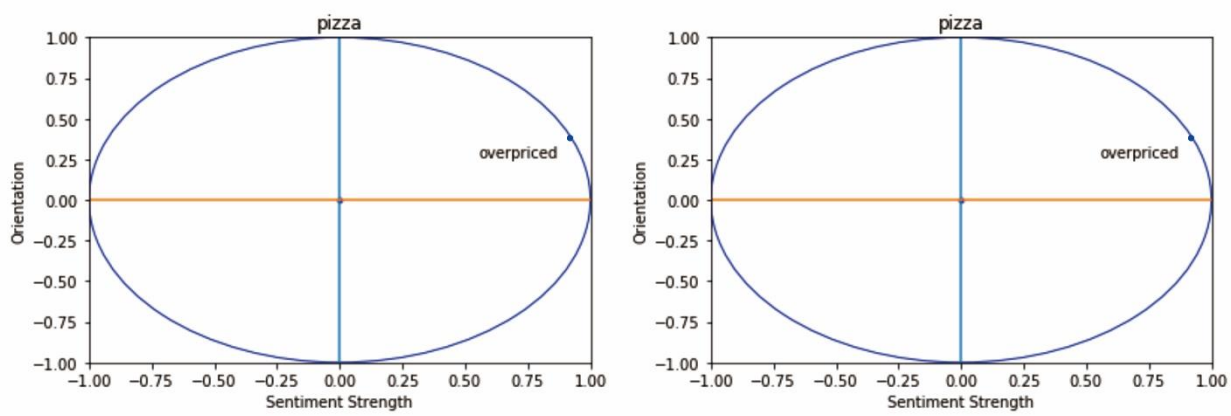
Review 4 result shows that SentiCircle in all performance ABSA1, ABSA2, and ABSA3 can extract properly the aspect and opinion word pair "food:satisfied" into the FOOD category of the aspect word "food" and the positive sentiment value of the opinion word "satisfied".

Review 5 result shows that SentiCircle on ABSA1 and ABSA2 cannot extract the aspect and opinion word pair "overpriced:overpriced" because they cannot recognize "overpriced" as the aspect word. Meanwhile, ABSA3 can produce the PRICE aspect category from the implicit aspect word "overpriced" that included in the implicit aspect lexicon. So, the opinion word "overpriced" that is generated from opinion lexicon can be extracted as the negative sentiment.

5. Conclusion and future work

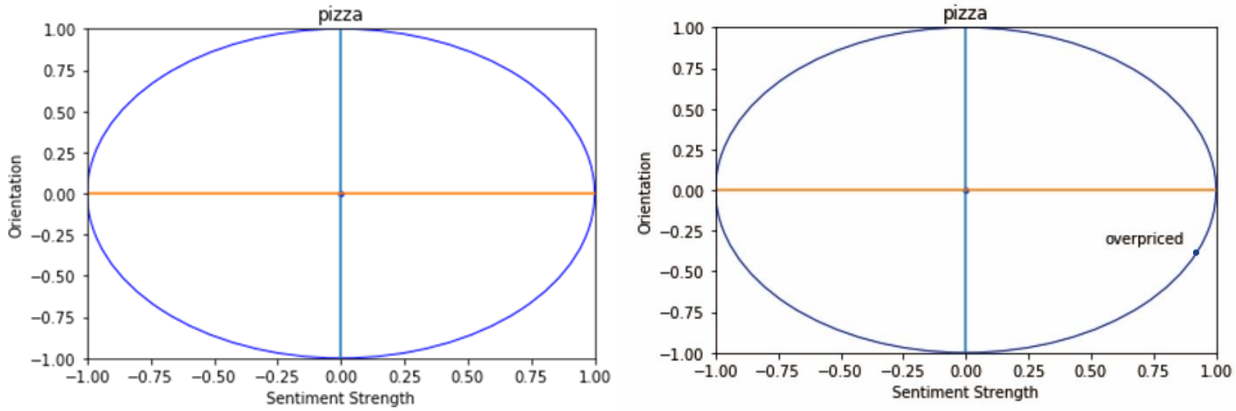
This research proposes ABSA for explicit and implicit aspects using grammatical rule extraction, hybrid approach, and SentiCircle.

Grammatical rule extraction and hybrid approach are used for aspect categorization. In aspect categorization stage, we can produce better for aspect and opinion words extraction than previous researchs. We do the extraction of explicit sentence cases properly using keywords from wikipedia.



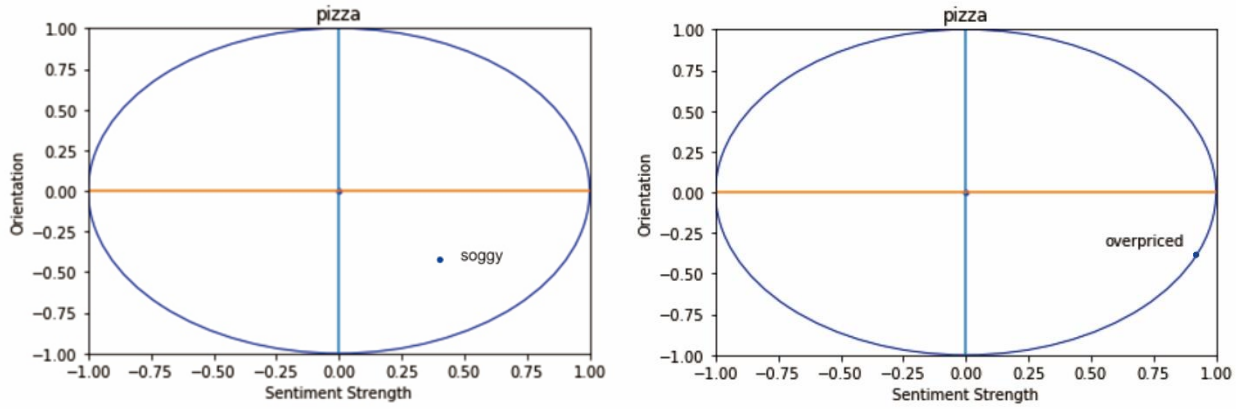
(a) (b)

Figure. 8 Review 2 extraction result by ABSA1: (a) in FOOD category and (b) in PRICE category



(a) (b)

Figure. 11 Review 2 extraction result by ABSA2: (a) in FOOD category and (b) in PRICE category



(a) (b)

Figure. 12 Review 2 extraction result by ABSA3: (a) in FOOD category and (b) in PRICE category

However, the wikipedia keyword does not work properly if it is used as a keyword for the extraction of implicit sentence cases. Here, we use wordnet to add the expansion of the aspect and opinion words that contained in the review based on the synonyms and antonyms of these words. We also add TF-ICF as an additional comparison for word similarity to increase the score accuracy of the aspect and opinion word that have been taken compared with the similarity of words in the review document. Furthermore, the results of the expansion of words from wordnet and TF-ICF are paired into the Wikipedia keyword to retrieve word similarities in sentences using Semantic Similarity. The result of the best aspect categorization performance is AC3 approach with F1-Measure of 0.86.

In aspect-based sentiment analysis process, we use SentiCircle, Opinion Lexicon, Implicit Aspect Lexicon, and Co-occurrence of context terms. SentiCircle changed the pairing of the word "overpriced" in review "the pizza was overpriced and soggy." which was originally paired with "pizza" and resulted in the FOOD category, then it was changed to produce the PRICE category according to the context of the sentence with positive sentiment polarity. Then by ABSA3 approach, the SentiCircle sentiment polarity is changed based on opinion lexicon, implicit aspect lexicon, and co-occurrence word. So, the result of the best aspect-based sentiment analysis performance is ABSA3 approach with F1-Measure of 0.89. ABSA3 expands the context word of SentiCircle lexicon using opinion lexicon, implicit aspect lexicon, and co-occurrence word.

For further research, the other methods of aspect categorization can be carried out to improve the accuracy of aspect and opinion words extraction, especially: 1) in other types of sentence structure, 2) in sentence case with implicit opinion, and 3) in sentence case with implicit aspect and opinion. In this research, we have not be able to work on the case of sentence extraction for aspect words with state initials too, such as: word “Chinatown” in the review "The rest of the dim sum, though pricey by Chinatown standards, is worth it." where we have not be able to take an aspect and opinion words pair "Chinatown standards:pricey". So we hope that further research can work on these similar problems too.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

This research can work well and successfully because of the following research contributions: Conceptualization by Rachmad Abdullah and Prof. Riyanarto Sarno; methodology and software by Rachmad Abdullah; validation, formal analysis, investigation, resources by Prof. Riyanarto Sarno, Suharyanto, and Rachmad Abdullah; data curation, writing-original draft preparation, and writing-review and editing by Rachmad Abdullah; visualization by Suharyanto; supervision by Prof Riyanarto Sarno; project administration and funding acquisition by Suharyanto.

Acknowledgments

This research was funded by the Indonesian Ministry of Education and Culture under Penelitian Terapan Unggulan Perguruan Tinggi (PTUPT) Program managed by Institut Teknologi Sepuluh Nopember (ITS) and under Riset Inovatif-Produktif (RISPRO) Invitation Program managed by Lembaga Pengelola Dana Pendidikan (LPDP).

References

[1] M. Hu and B. Liu, “Mining and summarizing customer reviews”, In: *Proc. of the 2004 ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD ’04*, p. 168, 2004.

[2] K. Schouten and F. Frasincar, “Survey on Aspect-Level Sentiment Analysis”, *IEEE Trans. Knowl. Data Eng.*, Vol. 28, No. 3, pp. 813-830, 2016.

[3] A. K Samha, Y. Li, and J. Zhang, “Apect-Based Opinion Extraction From Customer Reviews”, pp. 149-160, 2014.

[4] D. Khotimah and R. Sarno, “Sentiment Analysis of Hotel Aspect Using Probabilistic Latent Semantic Analysis, Word Embedding and LSTM”, *Int. J. Intell. Eng. Syst.*, Vol. 12, No. 4, pp. 275-290, 2019.

[5] R. Priyantina and R. Sarno, “Sentiment Analysis of Hotel Reviews Using Latent Dirichlet Allocation, Semantic Similarity and LSTM”, *Int. J. Intell. Eng. Syst.*, Vol. 12, No. 4, pp. 142-155, 2019.

[6] A. Firmanto and R. Sarno, “Aspect-based sentiment analysis using grammatical rules, word similarity and SentiCircle”, *Int. J. Intell. Eng. Syst.*, Vol. 12, No. 5, pp. 190-201, 2019.

[7] F. Nurifan, R. Sarno, and K. R. Sungkono, “Aspect Based Sentiment Analysis for Restaurant Reviews Using Hybrid ELMo-Wikipedia and Hybrid Expanded Opinion Lexicon-SentiCircle”, *Int. J. Intell. Eng. Syst.*, Vol. 12, No. 6, pp. 47-58, 2019.

[8] Y. Setiowati, “Pair Extraction of Aspect and Implicit Opinion Word based on its Co-occurrence in Corpus of Bahasa Indonesia”, pp. 73-78, 2019.

[9] I. Cruz, A. Gelbukh, and G. Sidorov, “Implicit Aspect Indicator Extraction for Aspect-based Opinion Mining”, *Int. J. Comput. Linguist. Appl.*, Vol. 5, No. 2, pp. 135-152, 2014.

[10] S. Gagić, D. Tešanović, and A. Jovičić, “The Vital Components of Restaurant Quality that Affect Guest Satisfaction”, *Turizam*, Vol. 17, No. 4, pp. 166-176, 2013.

[11] R. Panchendrarajan, M. N. N. Ahamed, B. Murugaiah, S. Prakhash, S. Ranathunga, and A. Pemasiri, "Implicit Aspect Detection in Restaurant Reviews," In: *Proc. of Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, pp. 128–136, 2016, doi: 10.18653/v1/w16-0421.

[12] W. H. Gomaa and A. A. Fahmy, "A Survey of Text Similarity Approaches", *Int. J. Comput. Appl.*, Vol. 68, No. 13, pp. 13-18, 2013.

[13] L. H. Suadaa and A. Purwarianti, "Combination of Latent Dirichlet Allocation (LDA) and Term Frequency-Inverse Cluster Frequency (TFxICF) in Indonesian text clustering with labeling", In: *Proc. of 2016 4th International Conference on Information and Communication Technology, ICoICT 2016*, pp. 1-6, 2016.

[14] M. E. Peters, M. Neumann, M. Iyyer, M. Gardner, C. Clark, K. Lee, and L. Zettlemoyer, "Deep Contextualized Word Representations," In: *Proc. of North American Association for Computational Linguistics*, pp. 2227–2237, 2018, doi: 10.18653/v1/n18-1202.

[15] M. Abadi, "TensorFlow: A system for large-scale machine learning", In: *Proc. of 12th USENIX Symp. Oper. Syst. Des. Implementation, OSDI 2016*, pp. 265-283, 2016.

[16] J. A. Hansen, E. K. Ringger, and K. D. Seppi, "Probabilistic explicit topic modeling using Wikipedia", *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, Vol. 8105 LNAI, pp. 69-82, 2013.

[17] C. Fellbaum, "WordNet(s)", in *Encyclopedia of Language & Linguistics*, pp. 665-670, 2006.

[18] E. G. Caldarola and A. M. Rinaldi, "Improving the visualization of word net large lexical database through semantic tag clouds", In: *Proc. of 2016 IEEE Int. Congr. Big Data, BigData Congr. 2016*, pp. 34-41, 2016.

[19] K. Kearns, *Semantics (Macmillan Modern Linguistics)*, Palgrave Macmillan, 2011.

[20] J. Martinez-Gil, "An overview of textual semantic similarity measures based on web intelligence", *Artif. Intell. Rev.*, Vol. 42, No. 4, pp. 935-943, 2014.

[21] P. Xia, L. Zhang, and F. Li, "Learning similarity with cosine similarity ensemble", *Inf. Sci. (Ny).*, Vol. 307, pp. 39-52, 2015.

[22] G. Qiu, B. Liu, J. Bu, and C. Chen, "Expanding domain sentiment lexicon through double propagation", *IJCAI Int. Jt. Conf. Artif. Intell.*, No. November 2014, pp. 1199-1204, 2009.

[23] H. Saif, M. Fernandez, Y. He, and H. Alani, "SentiCircles for contextual and conceptual semantic sentiment analysis of Twitter", *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2014.

[24] P. D. Turney and P. Pantel, "From frequency to meaning: Vector space models of semantics", *J. Artif. Intell. Res.*, Vol. 37, pp. 141-188, 2010.

[25] M. Pontiki, D. Galanis, H. Papageorgiou, S. Manandhar, and I. Androutsopoulos, "SemEval-2015 Task 12: Aspect Based Sentiment Analysis", 2015.

[26] S. Poria, E. Cambria, L. W. Ku, C. Gui, and A. Gelbukh, "A Rule-Based Approach to Aspect Extraction from Product Reviews", pp. 28-37, 2015.