Pair Extraction of Aspect and Implicit Opinion Word based on its Co-occurrence in Corpus of Bahasa Indonesia

Yuliana Setiowati
Department of Informatics, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia
yulianasetiowati.18051@mhs.its.ac.id

Arif Djunaidy
Department of Information Systems, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia
adjunaidy@is.its.ac.id

Daniel Oranova Siahaan
Department of Informatics, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia
danielos@cs.its.ac.id

***Abstract*—In implicit opinion sentences containing implicit opinions, the use of traditional dictionary-based method is no longer effective, since there is no any opinion word serving as a clue to identify the sentiment value. This study proposes an approach to obtain pairs of aspect and opinion word from sentences containing implicit opinion. Two methods are developed this approach. The first method serves to separate clauses from a compound implicit opinion sentence and refine its corresponding parse-tree from clauses as a result of the separation. The second method determines the opinion word from the implicit opinion clause. This method uses co-occurrence of aspect and opinion word based on a corpus of explicit opinion sentences. In this study, an initial experiment was conducted using a data set containing 30 implicit opinion sentences with 76 clauses. The proposed approach is capable of detecting pairs of aspect and opinion from the given corpus of Bahasa Indonesia.**

Keywords—implicit opinion sentence, co-occurrence, pair of aspect and opinion word, SPOPelK sentence structure

I. INTRODUCTION

The development of review is growing rapidly in the industry, including reviews of products, films, events, government policies, tourist attractions, and so on. It affects the development of user review using Bahasa Indonesia. User reviews can be found in Twitter [1][2], Facebook [3], and websites [4][5]. The benefits of user reviews are important; for example, a review of a product helps increase brand awareness, marketability, product development, and product quality assurance. Users can understand the advantages and disadvantages of the product. It can also be a form of a recommendation to other users prior to purchasing the product. For the government, reviews can help determine positive or negative responses of the community related to the implemented policies [6].

It may not be easy to determine the advantage and disadvantage aspects of a product, even though the information is written in the review. In user review, there is an indirect relationship among aspects, namely an aspect being close to other aspects or an aspect being the sub-aspect of others. Aspects being close to each other can be grouped into a new aspect, which is called aspect categorization. An aspect can become the sub-aspect of another, and a sub-aspect becomes the sub-aspect of another, and so on. The process of building the taxonomy of an aspect is also called knowledge [7]. Lack of knowledge in certain domains causes some researchers to extract knowledge from user review [7][8]. Knowledge-forming aspects of taxonomy can consist of one [9] or more levels [7]. In the process of building knowledge, it is important to extract and categorize aspects. The framework of this study is to build knowledge in certain domains; but prior to that, this study discusses the aspect extraction as a preparation to build the knowledge.

User review can be categorized as fact or opinion [10]. A fact is an objective expression of an entity, event, and property; while an opinion is usually a subjective expression which describes judgments or feelings about entities, events, and properties. Fact can be descriptive or user experience in using certain entities. Fact can also indirectly contain opinions, which is called implicit opinion; for example, "*Saya sangat kecewa karena proses check-in menunggu hampir dua jam*" (I am very disappointed because the check-in process took almost two hours), This fact contains an opinion with a negative sentiment value which has the same opinion as, "*Saya sangat kecewa karena proses check-in lama*" (I am very disappointed because the check-in process is long). Based on the aforementioned, opinion sentence can be redefined as a sentence which describes positive or negative opinion explicitly or implicitly [10]. Explicit opinion sentence is a subjective sentence which describes positive or negative opinion, while implicit opinion sentence is an objective sentence which contains positive or negative opinion implicitly.

Explicit opinion sentence is a subjective sentence which contains adjective and adverb. In Bahasa Indonesia, explicit opinion sentence is denoted not only by an adjective, but also by verb and noun [11]. In Bahasa Indonesia, some verbs and nouns are formed from the adjective, is called the word formation. Whereas in English there are not such a formation, it is hardly possible for a noun to be turned into verb; for example, the noun "book" and the verb "book" have different meanings.

Implicit opinion sentence is objective in terms of information obtained from real events that can be proven by accurate and precise data. The data can be quantitative (in the form of numbers) and qualitative (statements) [12]. As there is not any opinion word as a clue to identify the sentiment value, the traditional dictionary-based method is no longer effective. Moreover, it is necessary to determine the objective features for sentiment analysis of implicit opinion sentences. A number of researchers have identified implicit opinion sentences including Kang et al. [12] who developed a method to extract objective features based on two observations. Firstly, the objective feature is usually expressed by various lexicosyntactic structures, such as part-whole relations. Secondly, the objective feature is often stated quantitatively using number, weight, and size.

978-1-7281-4520-4/19/$31.00 ©2019 IEEE

If the implicit opinion sentence in Bahasa Indonesia has a predicate in the form of a verb, the verb will be neutral; for examples, "*terdengar*" (hear), "*menginap*" (stay), "*menyediakan*" (provide), "*menghadap*" (face), "*memesan*" (order), "*memberikan*" (give), etc. The predicate in the form of a verb can have a positive or negative sentiment value if it is supported by data. The data can be an object, complement, or adverb; example the implicit opinion sentence, "*proses check-in menunggu hampir dua jam*" (the process of checking-in took almost two hours) of which predicate "*menunggu*" (take) is neutral, supported by data "*dua jam*" (two hours), so the implicit opinion sentence has negative sentiment value. The opinion word contained in the implicit opinion sentence is "*menunggu lama*" (took a long time)

User review can use simple sentences, but there are users who tend to share their experience using complex sentences. In opinion sentence using English, simple sentence can be processed using Postagging [13] and complex opinion sentences can be processed using the complex NLP-based rules method [14] and dependency parsing [4][12]. In English, long sentences are generally referred to as complex sentences, while in Bahasa Indonesia they are called "*kalimat majemuk*" (compound sentence). Implicit opinion sentence is often contained in compound sentences. Unfortunately, however, the method used in complex opinion sentences in English cannot be directly implemented in Bahasa Indonesia.

The sentence structure in English has different grammatical rules from Bahasa Indonesia. English grammar has the NP and VP elements, while Bahasa Indonesia is arranged with the elements of "*Subjek*" (Subject), "*Predikat*" (Predicate), "*Objek*" (Object), "*Pelengkap*" (Complement) and "*Keterangan*" (Adverb) (SPOPelK). Based on the SPOPelK sentence structure, the opinion word and aspect occupy as one of the elements in the implicit opinion sentence using Bahasa Indonesia. Sibarani et al. [15] have established grammar rules for Bahasa Indonesia using NP and VP elements. However, the rules do not match the grammar of Bahasa Indonesia. Hence, this study develops parsing with the SPOPelK structure as the approach. The implicit opinion using compound sentence may consist of two or more clauses. Parsing with SPOPelK structure is used to separate clauses found in the implicit opinion sentence.

Zainuddin et al. [1] and Gojali et al. [16] conducted aspect extraction first and followed by determining pair of opinion word and aspect using the distance method. Regardless of the word elements in the grammar rules, this method allows the occurrence of errors in determining the pair of aspect and opinion words. On the other hand, Hai et al. [17] predict the implicit aspect of opinion sentence that only contain opinion word using the co-occurrence method. The result is in the form of a set of rules (knowledge) describing the relation between the opinion word with the aspect category. When this rule is implemented in a compound opinion sentence, the opinion word obtained will determine the right aspect or the aspect will determine the opinion word in the knowledge. The approach using the distance and co-occurrence methods is a statistical and predictive approach, thus allowing errors in determining the pair of aspect and opinion word. The process of parsing using the SPOPelK structure separates clauses; i.e if the clause is an explicit opinion, the aspect and opinion word will be obtained and if the clause is an implicit opinion, the supporting aspect and data will be obtained based on word structure in sentences.

To determine the sentiment value of an implicit opinion sentence, there is no any opinion word as a clue to identify the sentiment value, thus the traditional dictionary-based method is no longer effective. Liao et al. [22] stated that clauses in implicit opinion sentence generally have the same sentiment value according to the semantic context of the implicit opinion sentence. In a compound implicit opinion sentence, the sentiment value of the clause has the same or opposite sentiment value determined by the conjunction [23]. This study determines the positive or negative sentiment value by looking for the word opinion which corresponds to the meaning of a compound implicit opinion sentence.

This study proposes an approach to extract aspect and opinion word in implicit opinion sentences. Two methods were developed for this approach. The first method works for separating clauses from compound sentences and refining the resulting parse-tree from clauses as a result of the separation. The second method works to determine the opinion word from the clause. This method is based on the co-occurrence of aspect and opinion word in corpus containing opinion sentences.

The remaining of the paper is organized as follows. Section 2 outlines the related work. Section 3 explains the proposed methodology for extraction pair of aspect and opinion word in implicit opinion sentences in Bahasa Indonesia. Section 4 presents the experimental result. Section 5 concludes the paper.

II. Related Work

A. Compound Opinion Sentence

In this study, the implicit opinion sentence used is in the form of compound sentence (*kalimat majemuk*). The types of a compound sentence in Bahasa Indonesia consists of *setara* (equivalent), *bertingkat* (complex), and *campuran* (complex compound). Therefore, the opinion sentence used in this study is compound implicit opinion sentence [24].

Equivalent compound sentence (*kalimat majemuk setara*) is a compound sentence that consists of two sentence structures that have equivalent or parallel meanings. A single sentence (*kalimat tunggal/clause*) is a sentence which only consists of one core sentence or one clause. The essence of a sentence is formed from the subject, predicate, object and complement. Two sentences are combined into one sentence using conjunctions, such as "*dan*" (and), "*lalu*" (then), "*tetapi*" (but), and "*atau*" (or). The characteristic of equivalent compound sentences is that one sentence has a coordinative relationship with another sentence, so that it can be a stand-alone sentence even if they are separated. A complex sentence (*kalimat majemuk bertingkat*) is a sentence that has two sentence structures combined into one using subordinating conjunctions, such as "*jika*" (if), "*ketika*" (because), "*walaupun*" (when), "*meskipun*" (even though), "*karena*" (because). This study used complex implicit opinion sentences of equivalent compound sentences and complex sentences. The characteristic of a complex compound sentence is that one clause cannot be stand-alone; when the sentence is separated, since the clause will have no meaning. A complex compound sentence (*kalimat majemuk*

campuran) is a combination of equivalent compound and complex sentence; at least, it must consist of three simple sentences. For this context, consider the following two examplesof sentences containing implicit opinion.

1. Implicit opinion using equivalent compound sentence (*kalimat majemuk setara*)

Consider the sentence S1: "*Hotel berbintang bagus dan kualitas hotel tidak lekang oleh jaman*." (The hotel is amazing, and the hotel quality is timeless) that can be divided into two clauses, i.e. S1-1 (clause 1): "*Hotel berbintang bagus.*" (The hotel is amazing) and S1-2 (clause 2): "*Kualitas hotel tidak lekang oleh jaman.*" (Hotel quality is timeless).

The clause 1 is an explicit opinion sentence with a positive sentiment value, having the "hotel" aspect and the adjective "*bagus*" (amazing). The clause 2 is an implicit opinion sentence with a positive sentiment value, having the "*kualitas hotel*" (hotel quality) aspect and the supporting data "*tidak lekang oleh jaman*" (timeless).

2. Implicit opinion using complex sentence (*kalimat majemuk bertingkat*)

Consider the sentence S2: "*Hotel cukup nyaman walaupun terdengar suara bunyi alarm perlintasan kereta api pada saat malam hari.*" (The hotel is fairly comfortable even though I can hear the train siren at night) that can be divided into two clauses, i.e. S2-1 (clause 1): "*Hotel cukup nyaman*" (The hotel is fairly comfortable) and S2-2 (clause 2): "*Hotel kalau malam terdengar suara bunyi alarm perlintasan kereta api*" (I can hear the train siren at night);

The clause 1 is an explicit opinion sentence with a positive sentiment value, having the "hotel" aspect and the opinion word "*nyaman*" (comfortable). The clause 2 is an implicit opinion sentence with a negative sentiment value, having the "hotel" aspect and supporting data "*suara bunyi alarm perlintasan kereta api*" (the train siren).

This study focuses on compound implicit opinion sentences by: (1) separating implicit compound opinion sentence into clauses and (2) completing the clauses into a perfect sentence.

B. Implicit Opinion Sentence

Implicit opinion sentence is an objective sentence that contains positive or negative opinion implicitly. As there is no any opinion word as a clue to identify the sentiment value, the traditional dictionary-based method is no longer effective and it is necessary to determine the objective features for sentiment analysis of implicit opinion sentences. A number of researchers have identified implicit opinion sentences including Kang et al. [12] who developed a method to extract objective features based on two observations. First, the objective feature is usually expressed by various lexicosyntactic structures, such as part-whole relations. Second, the objective feature is often stated quantitatively using number, weight, and size.

In this study, it is necessary to identify the implicit opinion sentence and to determine the sentiment value of the implicit opinion sentence.

1. Identifying Implicit Opinion Sentence

This study uses implicit compound opinion sentences of which the clause contains neutral verbs, such as "*terdengar*" (hear), "*menginap*" (stay), "*menyediakan*" (provide), "*menghadap*" (face), "*memesan*" (order), "*memberi*" (give), etc. The predicate in the form of verb has a positive or negative sentiment value if it is supported by data. Data can be quantitative, in the form of numbers. The data can also be an object, complement, or adverb.

For example, the sentence S3: "*Kita harus menunggu mulai dari jam 2 sampai dengan jam 4*" (We have to wait from 2 pm to 4 pm) has the verb "menunggu" (wait) which is neutral and the adverb "*mulai dari jam 2 sampai dengan jam 4*" (from 2 pm to 4 pm) as supporting data that shows the length of time.

2. Determining Sentiment Value of Implicit Opinion Sentence

(a) In a compound opinion sentence, the sentiment value of the clause can have the same or opposite sentiment value based on the conjunction [22]. The conjunction "*dan*" (and) in the implicit opinion sentence has the same sentiment value as the sentiment value of the explicit opinion sentence, whereas the conjunction "*tetapi*" (but) has a different sentiment value than the sentiment value of the explicit opinion sentence.

(b) Relevance of semantic background takes part. The sentiment value of a fact is affected by the semantic context or background of the topic.

For example, consider the following three clauses, i.e.

S4-1: "*Stasiun kereta hanya berjarak 100 meter dari rumah saya*" (The train station is only 100 meters from my house); S4-2: "*Saya tidak khawatir terlambat kerja*" (I am not worried about being late for work); and S4-3: "*Sangat bising dan terlalu banyak lalu lintas setiap hari*" (It is very noisy and too much traffic every day).

The implicit opinion sentence S4-1 has a positive sentiment value because clause S4-2 constitute the benefits of "*rumah yang dekat dengan stasiun kereta api*" (the house being close to the train station). When the context of the semantic background changes, however, the sentiment value of the fact will change according to the background context. The implicit opinion sentence S4-1 may have a negative sentiment value due to the clause S4-3 constitutes the disadvantages of "*rumah yang dekat dengan stasiun kereta api*" (the house being close to the train station).

III. METHODOLOGY

This study proposes an approach to extract aspect and opinion word in implicit opinion sentences. Two methods were developed for this approach. The first method works for separating clauses from compound sentences and refining the resulting parse-tree from clauses as a result of the separation. The second method works to determine the opinion word from the clause. This method uses co-occurrence of aspect and opinion word based on corpus of opinion sentences. The overall system design is shown in Figure 1.

A. Determining Implicit Aspect and Supporting Phrase using Parsing of SPOPelK Structure

The data set is preprocessed manually by writing down the abbreviated words, replacing slang with formal form, change English words into Bahasa Indonesia. The next process is postagging and separating sentences based on

period mark "." This study uses parsing with SPOPelK structure to separate clauses in the implicit opinion sentence.

In the process of parsing, S denotes as a Sentence. A Sentence that can express compound sentences or clauses consists of subject, predicate, object, complement, and adverb. The subject in the clause can be a pronoun, noun, noun phrase, or nominal phrase. The predicate can be in the form of verb, adjective, or verb phrase. The object can be noun, noun phrase, or nominal phrase. The complement can be adjective, nominal phrase, or adjective phrase. The adverb can be prepositional phrases. An equivalent compound sentence that connects two clauses with conjunctions "dan" (and), "tetapi" (but), or "atau" (or) with the postagging CC will have "S CC S" grammar rules, as shown in formula 1. On the other hand, a complex sentence that connects two clauses with conjunctions "jika" (if), "karena" (because), or "ketika" (when) with the postagging SC will have "S SC S" grammar rules, as shown in formula 2. Both formula 1 and 2 are additional language rules for compound sentences.

$$S \rightarrow S\ CC\ S \quad (1)$$

$$S \rightarrow S\ SC\ S \quad (2)$$

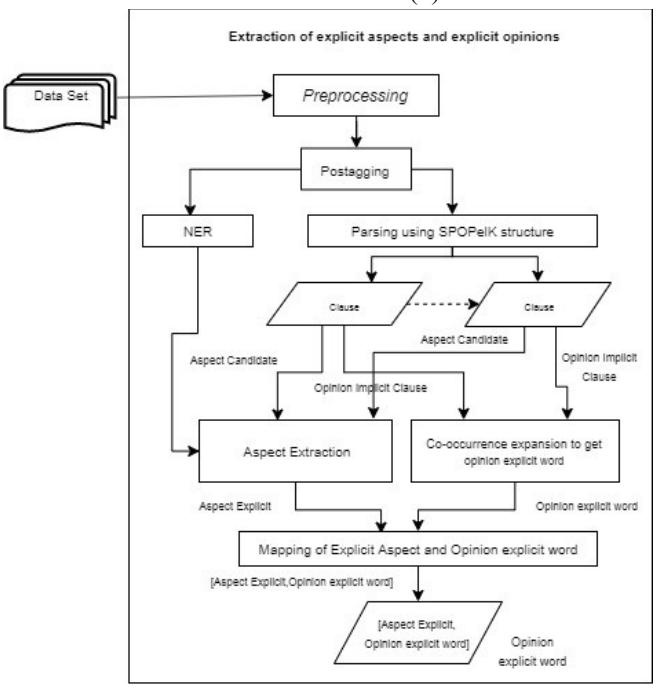

Fig 1. System Design

The process to obtain implicit and explicit aspects of compound opinion sentences is shown in Table I. The process is as follows.

1. The parsing process uses the *nltk.ChartParser* library with the parse() function. The initial parsing result is in the form of tree parsing and text;
2. Based on the parsing result, clauses are separated from the implicit compound opinion sentence. A clause is a sentence with the SPOPelK structure;
3. These clauses are determined whether or not they have met the SPOPelK structure. If not, it is completed into a perfect sentence in order to obtain the implicit aspect;
4. If the clause contains an explicit opinion, the candidate aspect and the opinion word can be obtained easily. If, otherwise, the clause contains an implicit opinion, the candidate aspect and supporting phrase are obtained.
5. Candidate aspect is pruned to obtain the valid aspect. The pruning process comprises similarity pruning, frequency pruning, and general concept.

B. Obtaining Opinion Word in Implicit Opinion Sentence using Co-occurrence Rules

An approach used to determine the opinion word that corresponds to the implicit opinion sentence is based on the co-occurrence rule. In this regard, the occurrence rule calculates the occurrence frequency of two words in one sentence. To build a co-occurrence matrix, training data are used in the form of corpus containing explicit opinion sentences. Each clause of these sentences may contain explicit or implicit opinion. An explicit opinion clause consists of an aspect (a_i) followed by an opinion word (o_i), while an implicit opinion clause consists of a bag of words (W_i). A co-occurrence matrix correlates between pairs of aspect (a_i) and opinion word (o_i) with a bag of words (W_{ik}) in the implicit opinion clause as shown in formula 3. In the co-occurrence matrix, a row represents an explicit opinion sentence, while the columns represent aspects (a_i), opinion words (o_i), and a bag of words of all implicit opinion clauses $W_i = \{W_{i1}, W_{i2}, \dots, W_{ik}, \dots, W_{in}\}$. If an explicit opinion sentence i contains the word W_{ik}, then a value of 1 is given in the column k, if not then a value of 0 is given. Furthermore, each word is searched for its synonyms using word2vec.

$$Explicit\ Sentence_i = a_i\ o_i, W_i;\\ and\ W_i = \{W_{i1}, W_{i2}, \dots, W_{ik}, \dots, W_{in}\} \quad (3)$$

Since the same or different sentiment values may exist in both explicit and implicit opinion clauses, the following three different rules are needed. Rule 1 handles both explicit and implicit opinion clauses that have the same sentiment values, while Rule 2 handles for both clauses that have different sentiment values. For an explicit sentence that contains several implicit opinion clauses with different sentiment values are handled by the Rule 3.

Rule 1: IF aspect = a AND opinion_word = o THEN W

This rule indicates that an explicit opinion sentence has clauses with the same sentiment value and the clause that contains implicit opinion does not contain negation.

For example, an implicit opinion sentence "*Saya sangat kecewa karena proses check-in menunggu hampir dua jam*" (I am very disappointed because the check-in process is long) can be detected using this rule, since it will fulfill the following condition:

Rule 2: IF aspect = a AND opinion_word = o THEN ~W
or equivalently
IF aspect = a AND opinion_word = ~o THEN W

This rule indicates that an explicit opinion sentence has clauses with different sentiment value, usually marked with the conjunction "only", "but"; and the clause that contains implicit opinion does not contain negation. This conjunction becomes the negation of the sentence of implicit opinion.

TABLE I. THE PROCES TO OBTAIN ASPECT AND SUPPORTING DATA

Opinion Implicit Sentence	Parsing Result using SPOPelK	Clauses Separating	Clauses Completing	Aspects and supporting data
hotel ini istimewa karena terdapat artistik sejarah (this hotel is special because there is artistic history)	(S (S (SUB (NNP (NN hotel) (PR ini))) (PRE (JJ istimewa))) (SC karena) (S (PRE (VB terdapat)) (OBJ (NNP (NN artistik) (NN sejarah)))))	**Clausa 1:** (S (SUB (NNP (NN hotel) (PR ini))) (PRE (JJ istimewa))) **conjuction :** (SC karena) **Clausa 2:** (S (PRE (VB terdapat)) (OBJ (NNP (NN artistik) (NN sejarah))))	**Clausa 1:** (S (SUB (NNP (NN hotel) (PR ini))) (PRE (JJ istimewa))) **conjuction :** (SC karena) **Clausa 2:** (S **(SUB (NNP (NN hotel) (PR ini)))** (PRE (VB terdapat)) (OBJ (NNP (NN artistik) (NN sejarah))))	**Clausa 1:** **Opinion Explisit Sentence** aspek : *hotel ini* (this hotel) kata opini : ***istimewa*** (special) **Clausa 2:** **Opinion Implicit Sentence** SUB : *hotel ini* (this hotel) PRE : *terdapat* KET : *artistik sejarah* (there is artistic history)

For example, consider the explicit opinion sentence is S5 containing two clauses, each of which correspond to explicit and implicit opinions. The conjunction of the two clauses is the word "cuma" (but), so the clauses has different sentiment value. The sentence S5-1 is equivalent to the sentence S5-2.

S5-1: "*Semua pelayanan* (a) *bagus* (o), *cuma saat check-in kami harus menunggu mulai dari jam 2 sampai dengan jam 4* (W_1)" . (All services (a) were good (o), but we had to wait for check-in from 2 pm to 4 pm (W1).)

S5-2: *Semua pelayanan* (a) *tidak bagus* ~(o), *saat check-in kami harus menunggu mulai dari jam 2 sampai dengan jam 4* (W_1). (All services (a) were not good ~ (o), we had to wait for check-in from 2 pm to 4 pm (W1).)

Rule 3: IF aspect = a AND opinion_word = o THEN $W_1 \sim W_2$
or equivalently
IF aspect = a AND opinion_word = o THEN $W_1 W_2$

This rule shows that an explicit opinion sentence contains three or more clauses, one clause with an explicit opinion, two or more clauses with an implicit opinion ($W_1, W_2, \ldots, W_n$). Some clauses with implicit opinions have negations. For example, consider the sentence S6 that meets the rule 3. In this regard, the sentence S6-1 is equivalent to the sentence S6-2.

S6-1: *Kami mendapatkan ruangan yang memiliki jendela* (W_1) *tapi tidak ada pemandangan sama sekali* (~W_2), *jadi kami kurang bisa menikmati* (o) *suasana Yogya dari Hotel* (a). (We stayed in room that has windows (W_1), but there was not any view at all (~W_2), so we could not enjoy (o) Yogya from the hotel (a).)

S6-2: *Kami mendapatkan ruangan yang memiliki jendela* (W_1) *dengan pemandangan pemandangan malioboro* (W_2), *jadi kami bisa menikmati* ~(o) *suasana Yogya dari Hotel* (a). (We stayed in room that has windows (W_1) with a view of Malioboro (W_2), so we could enjoy ~(o) Yogya from the hotel (a).)

IV. RESULTS AND DISCUSSION

For initial experiment, the proposed method for the aspect and word of opinion in the opinion sentence is implicitly tested and evaluated.

A. Data Set Preparation

For initial experiment, a data set of user reviews obtained from a hotel in Yogyakarta amounted to 30 reviews which, then, is broken down into 76 clauses are used. User reviews that were obtained from 2018 to 2019 are divided into five categories: service, room, food and beverage, facility and hotel, and surrounding. Table II shows the data used for experimental parts of aspect and implicit opinion.

TABLE II. DATA SET PREPARATION

No	Aspect Categorization	Implicit Opinion		Explicit Opinion		Type of aspects	Aspect	
		Pos	Neg	Pos	Neg		Explicit	Implicit
1	Service	0	6	3	1	6	9	1
2	Room	3	1	2	5	8	11	0
3	F &B	5	2	3	5	9	12	3
4	Facility	5	4	4	3	7	16	0
5	Hotel and Surrounding	10	1	6	1	6	12	6
	Total = 92.1%	23	14	18	15	36	60	10
		30.26%	18.42%	23.68%	19.74%		85.71%	14.29%

B. Evaluation of Implicit Aspect and Supporting Phrase using Parsing with SPOPelK Structure

Table III shows the evaluation results of the data set when using parsing with SPOPelK structure. One of the stages in the parsing process is to complete the clause to be a perfect sentence in order to determine its implicit aspect. However, if the clause is a complex sentence, the parsing cannot find the implicit aspect. Not all implicit aspects are found, for example Clause is a perfect sentence but has no aspect. This is shown by the experimental results that the accuracy of 79.99% and 84,99% were obtained for detecting implicit and explicit aspects, respectively; while the accuracy of 83.79% and 90.93% were obtained for identifying implicit and explicit opinion sentences, respectively.

TABLE III. EVALUATION RESULT OF IMPLICIT AND EXPLICIT OPINION USING SPOPELK PARSING

No	Aspect Categorization	Implicit Opinion		Explicit Opinion		Aspect	
		Pos	Neg	Pos	Neg	Explicit	Implicit
1	Service	4	2	3	2	6	1
2	Room	4	1	2	4	11	0
3	F & B	4	2	3	4	10	1
4	Facility	4	2	3	2	16	0
5	Hotel and Surrounding	8	0	6	1	10	5
	Total	24	7	17	13	54	8
	80.26%	**31.58%**	**9.21%**	**22.37%**	**17.11%**	**77.14%**	**11.43%**

C. Evaluation of the Use of Co-ocurence Rules for Obtaining an Opinion on Implicit Opinion Sentences

Table IV shows evaluation results of the use of co-occurrence rules for obtaining the aspect and opinion word on data set. Of 30 implicit opinion sentences used for experimental purpose, 26 sentences (86.67%) can be successfully identified by the co-occurrence rules. Among of these 26 implicit opinions, 50% of them met the co-occurrence rule1, 20% met the co-occurrence rule 2, and the remaining 16.67% met the co-occurrence rule 3. Large amount of training data, the co-occurrence matrix based on an aspect and the word opinion has a relationship with various implicit opinion sentences. It can increase the level of accuracy in obtaining the correct opinion word from the implicit opinion sentences.

TABLE IV. USE OF CO-OCCURRENCE RULES ON DATA SET

No	Aspect Categorization	Rule 1	2	3
1	Service	1	3	0
2	Room	3	0	1
3	F & B	4	0	1
4	Facility	3	0	2
5	Hotel and Surrounding	4	3	1
	Total	15	6	5
	86.67%	**50%**	**20%**	**16.67%**

V. CONCLUSION

An approach to obtain pairs of aspect and opinion word from sentences containing implicit opinion has been proposed in this study. The purpose of this study is to obtain implicit aspects and supporting data for implicit opinion sentences using the SPOPelK parsing process. Experimental results showed that the accuracy of 79.99% and 83.79% were obtained for identifying implicit aspects and implicit opinion sentences, respectively. Among 30 implicit opinion sentences used for experimental purpose, 26 sentences (86.67%) can be successfully identified by the co-occurrence rules, i.e. 50% of them met the co-occurrence rule1, 20% met the co-occurrence rule 2, and the remaining 16.67% met the co-occurrence rule 3. It can also be concluded that the larger the number of corpus data being used the higher the accuracy for finding opinion words that match the implicit opinion sentences. In further study, the larger amounts and different domains of data, such as hotel and restaurant, will be used to evaluate the proposed approach more thoroughly.

ACKNOWLEDGMENT

We would like to thank Institut Teknologi Sepuluh Nopember (ITS) as well as the Politeknik Elektronika Negeri Surabaya (PENS) that provided an opportunity to pursue my further study leading to a Doctoral degree.

REFERENCES

[1] N. Zainuddin, A. Selamat, and R. Ibrahim, "Hybrid sentiment classification on twitter aspect-based sentiment analysis," *Appl. Intell.*, vol. 48, no. 5, pp. 1218–1232, 2018.

[2] H. H. Lek and D. C. C. Poo, "Aspect-based Twitter sentiment classification," *Proc. - Int. Conf. Tools with Artif. Intell. ICTAI*, pp. 366–373, 2013.

[3] S. Setty, R. Jadit, and V. Mudenagudi, "Classification of Facebook News Feeds and Sentiment Analysis," in *International Conference on Advances in Computing, Communications and Informatics (ICACCI). doi:10.1109/icacci.201*, 2014, pp. 18–23.

[4] F. Lazhar and T. G. Yamina, "Mining explicit and implicit opinions from reviews," *Int. J. Data Mining, Model. Manag.*, vol. 8, no. 1, p. 75, 2016.

[5] W. Wang, H. Xu, and W. Wan, "Implicit feature identification via hybrid association rule mining," *Expert Syst. Appl.*, vol. 40, no. 9, pp. 3518–3531, 2013.

[6] Y. W. Syaifudin and D. Puspitasari, "Twitter Data Mining for Sentiment Analysis on Peoples Feedback Against Government Public," vol. 3, no. 1, pp. 110–122, 2017.

[7] A. Konjengbam, N. Dewangan, N. Kumar, and M. Singh, "Aspect ontology based review exploration," *Electron. Commer. Res. Appl.*, vol. 30, no. May, pp. 62–71, 2018.

[8] S. De Kok, L. Punt, R. Van Den Puttelaar, K. Ranta, K. Schouten, and F. Frasincar, "Review-Level Aspect-Based Sentiment Analysis Using an Ontology," 2018.

[9] N. Akhtar, N. Zubair, A. Kumar, and T. Ahmad, "Aspect based Sentiment Oriented Summarization of Hotel Reviews," *Procedia Comput. Sci.*, vol. 115, pp. 563–571, 2017.

[10] B. Liu, "Sentiment Analysis and Subjectivity," in *Handbook of Natural Language Processing*, Second Edi., Chapman and Hall/CRC, 2010, pp. 627–666.

[11] M. A. F. Yatim, Y. Wardhana, A. Kamal, A. A. R. Soroinda, F. Rachim, and M. I. Wonggo, "A corpus-based lexicon building in Indonesian political context through Indonesian online news media," *2016 Int. Conf. Adv. Comput. Sci. Inf. Syst. ICACSIS 2016*, pp. 347–352, 2017.

[12] Y. Kang and L. Zhou, "RubE: Rule-based methods for extracting product features from online consumer reviews," *Inf. Manag.*, vol. 54, no. 2, pp. 166–176, 2017.

[13] S. Li, L. Zhou, and Y. Li, "Improving aspect extraction by augmenting a frequency-based method with web-based similarity measures," *Inf. Process. Manag.*, vol. 51, no. 1, pp. 58–67, 2015.

[14] E. Marrese-Taylor, J. D. Velásquez, and F. Bravo-Marquez, "A novel deterministic approach for aspect-based opinion mining in tourism products reviews," *Expert Syst. Appl.*, vol. 41, no. 17, pp. 7764–7775, 2017.

[15] E. M. Sibarani, M. Nadial, E. Panggabean, and S. Meryana, "A Study of Parsing Process on Natural Language Processing in Bahasa Indonesia," *2013 IEEE 16th Int. Conf. Comput. Sci. Eng.*, pp. 309–316, 2013.

[16] S. Gojali and M. L. Khodra, "Aspect based sentiment analysis for review rating prediction," *4th IGNITE Conf. 2016 Int. Conf. Adv. Informatics Concepts, Theory Appl. ICAICTA 2016*, 2016.

[17] Z. Hai, K. Chang, and J. J. Kim, "Implicit feature identification via co-occurrence association rule mining," *Int. Conf. Intell. Text Process. Comput. Linguist.*, pp. 393–404, 2011.

[18] T. A. Rana and Y. N. Cheah, "A two-fold rule-based model for aspect extraction," *Expert Syst. Appl.*, vol. 89, pp. 273–285, 2017.

[19] D. Ekawati and M. L. Khodra, "Aspect-based sentiment analysis for Indonesian restaurant reviews," *Proc. - 2017 Int. Conf. Adv. Informatics Concepts, Theory Appl. ICAICTA 2017*, 2017.

[20] A. Candelieri and F. Archetti, "Detecting events and sentiment on Twitter for improving Urban Mobility."

[21] N. Dosoula, R. Griep, R. Den Ridder, R. Slangen, K. Schouten, and F. Frasincar, "Sentiment Analysis of Multiple Implicit Features per Sentence in Consumer Review Data," *Int. Balt. Conf. Databases Inf. Syst.*, pp. 289–303, 2016.

[22] J. Liao, S. Wang, and D. Li, "Identification of fact-implied implicit sentiment based on multi-level semantic fused representation," *Knowledge-Based Syst.*, vol. 165, pp. 197–207, 2019.

[23] D. H. Sasmita, A. F. Wicaksono, S. Louvan, and M. Adriani, "Unsupervised aspect-based sentiment analysis on Indonesian restaurant reviews," *Proc. 2017 Int. Conf. Asian Lang. Process. IALP 2017*, vol. 2018–Janua, pp. 383–386, 2018.

[24] T. Harsiati, A. Trianto, and E. Kosasih, *Buku Bahasa Indonesia SMP VII*. Pusat Kurikulum dan Perbukuan, Balitbang, Kemendikbud, 2017.

[25] A. Bagheri, M. Saraee, and F. De Jong, "Care more about customers: Unsupervised domain-independent aspect detection for sentiment analysis of customer reviews," *Knowledge-Based Syst.*, vol. 52, no. August, pp. 201–213, 2013.