WSD

by Setio B

Submission date: 05-Aug-2023 05:37AM (UTC+0700)
Submission ID: 2141448289
File name: aph_word_meaning_determination_by_LESK_algorithm_application.pdf (6.61M)
Word count: 5042
Character count: 30156

12th International Conference on Information & Communication Technology and System (ICTS) 2019

Word Sense Disambiguation (WSD) for Indonesian Homograph Word Meaning Determination by LESK Algorithm Application

Setio Basuki
Faculty of Engineering
Informatics Department
Universitas Muhammadiyah Malang
Indonesia, Malang
Email: setio_basuki@umm.ac.id

Ali Sofyan Kholimi
Faculty of Engineering
Informatics Department
Universitas Muhammadiyah Malang
Indonesia, Malang
kholimi@gmail.com

Agus Eko Minarno
Faculty of Engineering
Informatics Department
Universitas Muhammadiyah Malang
Indonesia, Malang
aguseko@umm.ac.id

Fauzi Dwi Setiawan Sumadi
Faculty of Engineering
Informatics Department
Universitas Muhammadiyah Malang
Indonesia, Malang
fauzisumadi@umm.ac.id

M. Rizal Arif Effendy
Faculty of Engineering
Informatics Department
Universitas Muhammadiyah Malang
Indonesia, Malang
muhammadrizalarifeffendy@gmail.com

***Abstract*— Indonesian has several words which are commonly known as ambiguous words, confusing the meaning of a sentence or a statement to be less understood or even not delivered. It is different to a human perception which has linguistic ability to determine the meaning of ambiguity or more than a word meaning. Word Sense Disambiguation is one of a topic from natural language processing (NLP) which deals with ambiguity handling. Word Sense Disambiguation is a linguistic computational process which aims to identify the proper meaning of words based on the context. This current study is designed as a system to handle the ambiguous words. It is conducted by looking up and defining the meaning of ambiguous words by using LESK algorithm. The test is performed towards the functionality from a system in which the result system test is in line with test data from KBBI. The result presents accuracy level of 78.6% for one of an ambiguous word and 62.5 % for two of ambiguous words in determining the meaning appropriately.**

Keywords*— *Word Sense Disambiguation, LESK Algorithm, Homograph, Ambiguity

I. INTRODUCTION

A language linguistic structure is a system which contains at least two matters. It is "utterance" and "meaning" [1]. Specifically, Keraf stated that word as a language vocabulary unit also contains of two aspects. Such aspect is known as "form" or "expressions" and "content" or "meaning" aspect [2] . In connection with the aforementioned definition of a language, it is mentioned that a language contains two aspects and one of them is meaning. In the language implementation, a language could have a word which has more than one meaning according to the succeeding sentences. The word that consists of more than one meaning could lead to a doubt in translating the meaning. This occurrence is commonly known as an ambiguity. However, in Indonesian, the problem of a word that has several meanings is not a new thing. The word that has more than a meaning is known as a homonym. Moreover, homonym itself is divided into two classifications, which are homophone and homograph. Homophone is a term which explains the two words or more having similar form in terms of its pronunciation and sound [3] for instance the words *bank* and *bang*. Those words have similar sound or pronunciation, but the meaning is different. Meanwhile, homograph is defined as the two words or more having similar form in term of its spelling or its writings [3]. For example, the word *teras* could have the meaning as 'the hard part of the woods' or 'space or floor in front of the house'.

The misinterpretation from the word meaning could lead the meaning of the sentence become unclear. The ambiguous word meaning of the homograph word group; therefore, becomes an exciting matter for the writer to be investigated. The homograph word group has the similar form of spelling and writing. Thus, it is common that people will have some difficulties to translate and to interpret the meaning according to the sentence context. Besides, human could differentiate the meaning of a word easily due to their linguistic ability, which is an ability to employ the words effectively either it is spoken or written. The sensitivity towards a sound, structure, meaning, and word function of a language is also possessed by human. Such ability is not possessed by a machine (computer). Therefore, the writer is encouraged to build an autonomous word meaning translation system (homograph) to assist the understanding about ambiguity of a word.

Computer technology world has a knowledge branch which studies about ambiguous word meaning handling. The ambiguous word meaning handling is included at one of Natural Language Processing (NLP) topics which is known as Word Sense Disambiguation (WSD) [4]. According to [5], Word Sense Disambiguation is a Computational linguistic process which is applied to identify the word proper meaning according to its context. WSD is often applied integrated in an application such as translation engine, news/information extractor, Q&A engine, opinion summary engine, and others. There are several researches that discuss about Word Sense Disambiguation. For instance, [6], enclosed inside a research, proposing a graph forming concept for word labelling in an English sentence. Thus, the most appropriate word meaning with the given context is able to be recognized. Moreover, in the second research [7] in the Word Sense Disambiguation topic, the writer explains that the utilization of LESK algorithm has several disadvantages. One of them is the applied source which is a normal dictionary, having a short definition. In order to treat this condition, the writer adapts the LESK algorithm which is supported by KBBI and WordNet. WordNet could be described as a dictionary which is similar to a normal dictionary, but it has more advantages in term of semantic relation [8,14,15]. In addition, this finding presents

978-1-7281-2133-8/19/$31.00 ©2019 IEEE

the fact that LESK algorithm achieves an 83% accuracy increment compared to a normal dictionary data system in the two researches in which the language is in English. In the progression of this Final Task, the writer attempts to apply Word Sense Disambiguation topic in Indonesian language. This effort is divided into chronological steps and process compared to English sentences. In addition, to support semantic database which is provided by WordNet, the writer also supplies homograph data along with its definition and sentence example from *Kamus Besar Bahasa Indonesia* (KBBI) or Indonesia Dictionary. Furthermore, there were two former researches that applied the LESK Algorithm for finding the ambiguous words including [9] and [10] in Hindi and Sinhala language. The authors from both papers successfully investigated the precision of the specified method which pointed approximately about 63% indicating that the algorithm could disambiguate words efficiently. Along with the growth of the context window of the tested data, the overall accuracy was rise accordingly. From the problems and several aforementioned research discussions, the current research emphasizes on defining the meaning of homograph word group automatically. The steps are performed in phases comprising of: pre-processing data, homograph word look up, and appropriate meaning word selection with the highest score from the application of LESK algorithm.

II. RESEARCH METHOD

The utilized method to accomplish this current research is included into one of Natural Language Processing (NLP) programs known as Word Sense Disambiguation. The employed algorithm is LESK algorithm, which is based on intuition where ambiguous word exists in a sentence together. It is used to refer to the similar topic and related meaning with the defined topic in the dictionary by using the similar word.

The LESK algorithm has relatedness function which will return the number of overlapping words between two inputted words. The following figure is a pseudo code of LESK algorithm [11].

```
for every word w[i] in the phrase
        let BEST_SCORE = 0
        let BEST_SENSE = null
        for every sense sense[i] of w[i]
                let SCORE = 0
                for every other word w[k] in the phrase, k != i
                        for every sense sense[l] of w[k]
                                SCORE = SCORE + number of words that occur in the glosss
                                of both sense[i] and sense[l]
                        end for
                end for
                if SCORE > BEST_SCORE
                        BEST_SCORE = SCORE
                        BEST_SENSE = w[i]
                end if
        end for
        if BEST_SCORE > 0
                output BEST_SENSE
        else
                output "Could not disambiguate w[i]"
        end if
end for
```

Figure 1. LESK Algorithm Pseudo Code

Figure 2 explains several processes/ steps which exist in the created system covering: pre-processing data, homograph word look-up, overlap score counting, and the highest score determination. The system is drawn generally in the following chart.

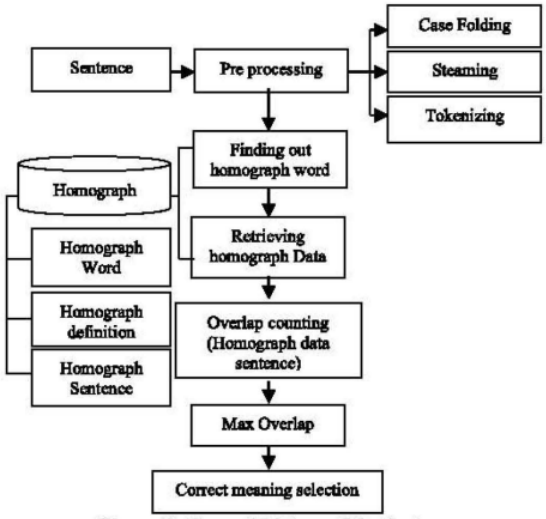

Figure 2. General Picture of the System

A. *LESK Algorithm*

LESK algorithm is based on the intuition where the ambiguous words exist together in a sentence and is used to refer the similar topic and related meaning. Based on the Figure 2, the LESK algorithm was performed on four main process including finding the homograph word, retrieving the homograph data, calculating the overlapping score, and pointing the highest score. The topic is defined in the dictionary by using similar word [12]. LESK algorithm has relatedness function that will return the number of overlapping words between the two inputted words definition [13]. The following is an example about the LESK algorithm implementation:

- Inputted words,

Apabila tidak puas dengan putusan pengadilan negeri boleh minta apel kepada pengadilan tinggi If not satisfied with the decision of the district court, one may request an appeal to the high court

Retrieve every appropriate meaning according to the homograph word found. At the input above, it is found that the word "*apel*" is in homograph database; therefore, every data meaning from the word "*apel*" is displayed in Table 1. Then, it is continued with tallying the words or looking up for the overlap in every word which is mutual with LESK algorithm concept.

TABLE I. *APEL* WORD MEANING AND OVERLAP LOOK-UP

apel			
Meaning 1	**Meaning 2**	**Meaning 3**	**Meaning 4**
Wajib hadir dalam suatu upacara resmi sifat militer tahu dengar amanat bendera naik	*Pohon buah bundar daging tebal kandung air kulit lunak warna merah kuning*	*Naik banding adil lebih tinggi mohon periksa ulang tingkat hadap putus pertama negeri*	*Kepala kampung* Chief of village

turun hormat bangsa besar kumpul banyak perintah paripurna lengkap ikut seluruh anggota siaga kesiap laksana tugas Must attend an official military ceremony To know why the flag goes up and down To respect as a great nation To get together with in plenary all members followed orders by standing by to handle the task	*matang rasa manis masam pyrus malus* A tree with round flesh and thick fruit with soft skin water contain when ripe it has sweet and sour taste a family of *Pyrus malus*	Appeal for higher justice double-check the level of the state court first verdict	
Meaning ∩ Inputted Sentence			
*apabila puas **putus adil negeri** minta **tinggi*** if satisfied with the decision, the district court ask for higher appeal			

TABLE II. *OVERLAP CALCULATION*

Overlap score of each meaning (Meaning ∩ Inputted Sentence)			
Meaning 1	**Meaning 2**	**Meaning 3**	**Meaning 4**
0	0	4	0

After the overlap calculation is done, the step is continued to the homograph words to determine which meaning has the highest score such as in Table 2. From the four meanings of *Apel*, it is retrieved that the highest score is in the third meaning. Therefore,

- *Apel* word in sentence, is:

> *Apabila tidak puas dengan putusan pengadilan negeri boleh minta **apel** kepada pengadilan tinggi.*
>
> If not satisfied with the decision of the district court, requesting an appeal from the high court is allowed.

- Has the meaning:

> *(nomina), naik banding pada pengadilan yang lebih tinggi; permohonan pemeriksaan ulang pada pengadilan tingkat kedua (pengadilan tinggi) terhadap keputusan pengadilan tingkat pertama (pengadilan negeri)*
>
> (Noun), appeal to a higher court; request for re-examination at the second level court (high court) towards the decision of the first court (district court)

B. Research Data

The data in this research uses homonym data (including homograph data), as many as 78 homograph and homonym words, 230 sentences that contains homograph and homonym words data, 162 meanings from the homograph and homonym words as enclosed in Table 3. All of the words, sentences and meanings from the homograph and homonym are derived from *Kamus Besar Bahasa Indonesia* (KBBI) and WordNet Bahasa. The homograph and homonym will be saved to the MySQL database. The following table shows the used homograph data example.

TABLE III. *HOMOGRAPH DATA*

III. RESEARCH FINDINGS AND DISCUSSION

The test is done by examining 140 Indonesian standard sentences which are obtained from KBBI and 30 end-users

No	Homograph word	Word class	Meaning	Sentence sample
1.	*Bisa*	noun	*zat racun yg dapat menyebabkan luka, busuk, atau mati bagi sesuatu yg hidup (biasanya terdapat pd binatang)* toxic substances that can cause injury, decay, or die for (usually in animals)	*Bisa ular sangat berbahaya* Snake venom can be dangerous
		Verb	*mampu melakukan sesuatu, dapat* Capable of doing something	Dia *bisa* membaca tapi tidak dengan menulis He can read but can't write

with the details of 70 sentences from KBBI and 70 sentences from end-users. This test is projected to know whether the meaning at defining system of a homograph word produces the mutual result with the homograph word meaning. Table 4 presents the test result towards 140 Indonesian standard sentences.

TABLE IV. *ONE AMBIGUOUS WORLD TEST DATA*

No	Tested Sentence	Relevant/ Irrelevant
1	*aku membaca **buku** tentang sejarah di perpustakaan* **(I read history book in the library)**	*Sesuai* (Relevant)
2	*Tebu itu mempunyai banyak **buku*** **(Cane has much segment)**	*Sesuai* (Relevant)
3	*pada **teras** di bawahnya terbentang taman yang luas* **(A wide garden lies under the terrace)**	*Sesuai* (Relevant)
4	*pada irisan melintang kelihatan **teras** yang dikelilingi lingkaran-lingkaran kayu yang dapat menunjukkan umurnya* **(on the slice of the cross section, a porch is surrounded by wooden circles that can indicate its age)**	*Sesuai* (Relevant)
5	*bukan hanya pembangunan fisik yang diperhatikan, melainkan juga pembangunan **mental*** **(not only is physical development to be considered, but also is mental development)**	*Sesuai* (Relevant)
6	*akibat tabrakan yang terjadi tubuh itu **mental** sekitar 2 meter* **(due to a collision that occurred, the body was bouncing about 2 meters)**	*tidak sesuai* (Irrelevant)
7	*perkara mesin, dia lebih **tahu** daripada saya* **(Dealing with machines, he knows better than me)**	*tidak sesuai* (Irrelevant)
8	*ayah membawa **tahu** sumedang saat pergi ke sumedang* **(fathers brought Sumedang tofu when he went to Sumedang)**	*Sesuai* (Relevant)
9	*padi jangan ditanam terlalu **rapat*** **(rice should not be planted too tightly)**	*tidak sesuai* (Irrelevant)
10	*kami mengadakan **rapat** di dalam ruangan* **(we hold meetings in the room)**	*Sesuai* (Relevant)
11	*matahari terbenam menandakan **malam** telah tiba* **(the sunset signifies that the night has come)**	*Sesuai* (Relevant)
12	*ia kehabisan **malam** saat membatik* **(he ran out of ink when making batik)**	*Sesuai* (Relevant)
13	*dia mendapatkan **bunga** mawar merah yang harum dari penggemarnya* **(he gets fragrant red roses from his fan)**	*Sesuai* (Relevant)
14	*dia mendapatkan **bunga** bank atas jasa investasi modalnya* **(he gets a bank interest for his capital investment services)**	*Sesuai* (Relevant)
15	*sepatu dengan **hak** tinggi sedang digemari oleh wanita karier* **(shoe with high heel is favored by career women)**	*Sesuai* (Relevant)
16	*ia juga punya **hak** tinggal di sini* **(he also has the rights to live here)**	*Sesuai* (Relevant)
17	*satu minggu ini, dia sudah empat **kali** datang ke rumahku* **(for this week long, he has come to my house four times)**	*Sesuai* (Relevant)
18	*ia melihat bebek berenang di **kali*** **(he sees a duck swimming in the river)**	*tidak sesuai* (Irrelevant)
19	*ia mendapat **gelar** srikandi dari kawan-kawannya* **(he received the title of *srikandi* from his friends)**	*Sesuai* (Relevant)
20	*tolong gelar **tikar** itu* **(please fold that mat)**	*Sesuai* (Relevant)
21	***jarak** antara Mekah dan Medinah kami tempuh dengan bus dalam 5 jam* **(we traveled distances between Mecca and Medina by bus in 5 hours)**	*Sesuai* (Relevant)
22	*ia sedang mencari biji **jarak*** **(he is looking for castor seed)**	*Sesuai* (Relevant)
23	*ia mendapat **salam** dari ayahnya* **(he received greeting from his father)**	*Sesuai* (Relevant)
24	*ia lupa memasukkan **salam** kemasakannya* **(he forgot to give *salam* leave in his cooking)**	*Sesuai* (Relevant)
25	*hal ini sudah **jamak** terjadi* **(this has happened so often)**	*tidak sesuai* (Irrelevant)
26	*ia sedang men**jamak** salat* **(he is grouping his prayers)**	*Sesuai* (Relevant)
27	***roman** wajahnya berubah menjadi sedih* **(he feature of his face turned sad)**	*Sesuai* (Relevant)
28	***roman** lebih banyak membawa sifat-sifat zamannya drama atau puisi* **(romances carry more of the characteristics of the era of drama or poetry)**	*Sesuai* (Relevant)
29	*buka **buku** pada halaman 50* **(open the book on page 50)**	*Sesuai* (Relevant)
30	***halaman** rumahnya ditanami cemara* **(Yard in his house is planted with pine trees)**	*Sesuai* (Relevant)
31	*angin terjadi karena gerakan **hawa*** **(winds occur due to air movement)**	*Sesuai* (Relevant)
32	*ia tidak dapat menahan **hawa** nafsunya* **(he cannot resist his lusts)**	*Sesuai* (Relevant)
33	*baik sanak keluarga maupun orang **helat** banyak yang hadir pada resepsi pernikahannya* **(Many relatives and other guests were present at the wedding reception)**	*Sesuai* (Relevant)
34	*banyak tamu yang datang untuk meramaikan **helat** putri tunggalnya* **(Many guests came to enliven the strands of their only daughter)**	*Sesuai* (Relevant)
35	***hemat** pangkal kaya, rajin pangkal pandai* **(saving is the basis of being rich, diligent is the basis of being clever)**	*Sesuai* (Relevant)
36	*ia mendengarkan pelajaran dengan **hemat** dan cermat* **(he listens to the lesson attentively and carefully)**	*tidak sesuai* (Irrelevant)

37	*majikan itu sangat **kejam**, tidak mau menaikkan upah buruhnya barang sedikit juga* **(the employer was very cruel that he did not want to raise the wages of his laborers, even in a little)**	*Sesuai* (Relevant)
38	*sungguhpun matanya **kejam**, ia tidak tidur* **(though his eyes were drowsy, he did not sleep)**	*Sesuai* (Relevant)
39	*perkataannya **lemak** manis* **(his word is awesomely sweet)**	*tidak sesuai* (Irrelevant)
40	***lemak** penyelar daging* **(fat bolstering in meat)**	*Sesuai* (Relevant)
41	*berbagai macam dampak **negatif** dapat disebabkan oleh sejarah politik* **(various kinds of negative impacts can be caused by political history)**	*tidak sesuai* (Irrelevant)
42	*kabel positif jangan langsung kau hubungkan dengan kabel **negatif*** **(positive cables must not be connected directly with the negative cable)**	*Sesuai* (Relevant)
43	***pupus** sudah harapannya selama ini* **(his hopes have been lost so far)**	*tidak sesuai* (Irrelevant)
44	***pupus** daun pisang tersebut terjatuh* **(the foliage of the banana leaves fell out)**	*Sesuai* (Relevant)
45	*surat tersebut dikirim melalui kantor **pos*** **(letter was sent through the post office)**	*Sesuai* (Relevant)
46	***pos** tentara militer tepat berada di depan* **(military army post is right in front of us)**	*Sesuai* (Relevant)
47	*di negara yang **rusuh** itu sering timbul pemberontakan* **(In a riotous country, rebellions often arise)**	*Sesuai* (Relevant)
48	*pada **babak** pertama, kedua kesebelasan masih bermain sama kuat* **(in the first round, both teams still play equally strong)**	*Sesuai* (Relevant)
49	*roni **babak** belur dihajar oleh lawannya* **(roni is battered as beaten by his opponent)**	*Sesuai* (Relevant)
50	*Ular kobra memiliki **bisa** yang sangat berbahaya dan mematikan* **(Cobra snake has a very dangerous and lethal venom)**	*Sesuai* (Relevant)
51	*Dia **bisa** membaca tapi tidak dengan menulis* **(He can read but cannot write)**	*tidak sesuai* (Irrelevant)
52	*buah **apel** sangat kaya akan vitamin C* **(apple isvery rich in vitamin C)**	*Sesuai* (Relevant)
53	*apabila tidak puas dengan putusan pengadilan negeri boleh minta **apel** kepada pengadilan tinggi* **(if unsatisfied with the verdict of the district court, an appeal from the high court might be requested)**	*Sesuai* (Relevant)
54	*amanat **apel** pagi dibacakan oleh pemimpin apel* **(morning ceremony preach is read by the leader of the ceremony)**	*Sesuai* (Relevant)
55	***Genting** rumahku rusak akibat angin puting beliung kemarin malam* **(Roof of my house was damaged by a tornado last night)**	*Sesuai* (Relevant)
56	*setelah perundingan menemui jalan buntu, keadaan bertambah **genting*** **(after negotiations have reached a dead end, the precarious situation is getting tensed)**	*tidak sesuai* (Irrelevant)
57	*dia merasakan sesak akibat **asma** yang dideritanya* **(he felt the tightness caused by his asthma)**	*Sesuai* (Relevant)
58	*kami berbuat baik terhadap siapapun semata-mata untuk meluhurkan **asma** Tuhan* **(we do good to anyone solely in the name of God)**	*Sesuai* (Relevant)
59	*semua orang heran bahwa istrinya dapat bersikap baik pada **madu**nya* **(everyone is surprised that his wife can be nice to his mistress)**	*tidak sesuai* (Irrelevant)
60	*sarang lebah ini jika diperas mengeluarkan **madu*** **(this beehive bears honey if squeezed)**	*Sesuai* (Relevant)
61	*tamu itu **gondok** karena diperlakukan tidak sewajarnya* **(the guests were mad because they were treated inappropriately)**	*tidak sesuai* (Irrelevant)
62	***gondok** merupakan penyakit pembengkakan pada leher* **(goiter is a swelling of the neck)**	*Sesuai* (Relevant)
63	*ayah sedang membaca cerita **bajak** laut* **(father is reading pirate story)**	*Sesuai* (Relevant)
64	***bajak** selalu di tanah yang lembut* **(always plow on the loose ground)**	*Sesuai* (Relevant)
65	***ahli** waris anggota keluarga tersebut adalah teman kerjaku* **(that family member heirs is my colleagues)**	*Sesuai* (Relevant)
66	*dia seorang yang **ahli** menjalankan mesin itu* **(he is an expert in running the machine)**	*Sesuai* (Relevant)
67	*tiang itu rusak dimakan **bubuk*** **(that poles was damaged and eaten by insect)**	*Sesuai* (Relevant)
68	***bubuk** kopi tersebut adalah yang paling hitam dan halus* **(that coffee powder is the blackest and the smoothest)**	*Sesuai* (Relevant)
69	*sabun yang baik banyak mengandung **busa*** **(good soap contains a lot of foam)**	*Sesuai* (Relevant)
70	***busa** jok mobil ini sangat empuk* **(this car seat foam is very soft)**	*Sesuai* (Relevant)
71	***Dasar** anak tidak berbakti!* **(What a filial children!)**	*Tidak Sesuai* (Irrelevant)
72	*Buah **apel** batu sangat segar jika di jus* **(*Batu* apples are very fresh when juiced)**	*Sesuai* (Relevant)
73	***Apel** kebangsaan pagi di lapangan tersebut dihadiri*	*Sesuai* (Relevant)

	seluruh pns kota malang **(Morning nationality ceremony on the field were attended by all Malang city civil servants)**	
74	***Tahu** asli kediri sangat gemar dijadikan oleh oleh khas* **(Kediri original tofu is very appropriate to be used as the typical gift)**	*Sesuai* (Relevant)
75	*Dia **tahu** sifat asli temannya yang di kira jahat tersebut* **(He knows the true nature of his friend who is thought to be evil)**	*Tidak Sesuai* (Irrelevant)
76	*toilet untuk kaum **hawa** pada tahap perbaikan* **(toilets for women are under the repair phase)**	*Tidak Sesuai* (Irrelevant)
77	***hawa** pada pagi hari ini terasa sejuk sekali* **(this morning feels very fresh)**	*Sesuai* (Relevant)
78	*gunakan **jangka** sebagai alat bantu untuk membuat lingkaran* **(use a bow as a tool to make circle)**	*Sesuai* (Relevant)
79	*menuju **teras** rumah hanya butuh beberapa langkah dari pintu* **(to reach the porch of the house, it only took a few steps from the door)**	*Sesuai* (Relevant)
80	*udara **malam** ini tidak begitu dingin* **(the night air is not so cold)**	*Tidak Sesuai* (Irrelevant)
81	***bunga** di pagi hari dibasahi oleh embun pagi* **(flower in the morning is moistened with morning dew)**	*Sesuai* (Relevant)
82	*dia rela di **madu** oleh suaminya* **(she is willing to be cheated by her husband)**	*Sesuai* (Relevant)
83	*setiap balita punya **hak** untuk imunisasi* **(every toddler has the rights to immunization)**	*Sesuai* (Relevant)
84	*jangan menggunakan sepatu **hak** tinggi dalam prosesi wisuda* **(do not stand on high heel shoe in the graduation process)**	*Sesuai* (Relevant)
85	*martabak **asal** kota malang memiliki cita rasa yang nikmat* **(*Martabak* from Malang city has a delicious taste**	*Sesuai* (Relevant)
86	***babak** pertama kontes take me out di menangkan oleh effendy* **(first round of take me out contest is won by Effendy)**	*Tidak Sesuai* (Irrelevant)
87	*risqi aris **babak** belur dihajar oleh tetangganya* **(Risqi Aris was battered and beaten by his neighbor)**	*Sesuai* (Relevant)
88	*Selai rasa **apel** memang paling banyak disukai anak-anak* **(Apple flavor jam is the most preferred by children)**	*Tidak Sesuai* (Irrelevant)
89	*Pertandingan sepak bola kemarin menghasilkan skor **seri*** **(Yesterday's soccer match was draw)**	*Tidak Sesuai* (Irrelevant)
90	***Tahu** makanan kesukaanku* **(Tofu is my favorite food)**	*Sesuai* (Relevant)
91	*Istri mana yang mau di **Madu***	*Sesuai*
	(Which wife wants to be cheated)	(Relevant)
92	*Setiap tahun, ani menerima **bunga** 5% dari bank* **(Every year, Ani receives 5% interest from the bank)**	*Sesuai* (Relevant)
93	***Bunga** mawar di taman tampak layu.* **(Roses in the garden look wilted)**	*Sesuai* (Relevant)
94	***Madu** mempunyai banyak manfaat untuk kesehatan tubuh.* **(Honey has many benefits for body health)**	*Sesuai* (Relevant)
95	*Umumnya semua istri tidak akan pernah mau di**madu**.* **(Generally, all wives will never want to be cheated)**	*Sesuai* (Relevant)
96	*Saat kau ingin rumahmu terang alami, pakailah **genting** kaca di bagian tertentu atap rumahmu* **(When you want your house to be naturally bright, wear a glass tile on a certain part of your roof)**	*Sesuai* (Relevant)
97	*Keadaan bertambah **genting** mendengar kabar ani diculik.* **(The situation grew precarious when hearing that Ani was kidnapped)**	*Tidak Sesuai* (Irrelevant)
98	*Aku ingin menjadi raja **bajak** laut* **(I want to be a pirate king)**	*Sesuai* (Relevant)
99	*Zoro mem**bajak** sawah* **(Zoro plows the fields)**	*Sesuai* (Relevant)
100	***Halaman** yang indah menjadi idaman setiap org yg punya rumah* **(Beautiful garden is the dream of every person who has a house)**	*Sesuai* (Relevant)
101	*Bagian HRM mau mengadakan **rapat** nanti sore* **(The human resource management section wants to hold a meeting later in the evening)**	*Sesuai* (Relevant)
102	*kita dulu berada pada **rahim** ibu* **(we used to be in the mother's womb)**	*Tidak Sesuai* (Irrelevant)
103	*tahun 2015 Cassilas **hengkang** dari real madrid* **(In 2015, Cassilas left Real Madrid)**	*Tidak Sesuai* (Irrelevant)
104	*hari senin terjadi **baku** hantam antar siswa smp di Malang* **(On Monday, there were fights between junior high school students in Malang)**	*Sesuai* (Relevant)
105	*pencuri memasuki rumah dengan memanjat **genting*** **(thieves entered the house by climbing roof)**	*Sesuai* (Relevant)
106	*menikmati kopi di **halaman** depan rumah* **(enjoying coffee in the front yard of the house)**	*Sesuai* (Relevant)
107	*Para guru hadir pagi hari untuk melakukan **apel*** **(Teachers present in the morning to do ceremony)**	*Sesuai* (Relevant)
108	***Apel** itu banyak yang busuk.* **(The apples are rotten a lot)**	*Tidak Sesuai* (Irrelevant)
109	*Kakak selalu bersantai setiap sore di **teras** rumah*	*Sesuai* (Relevant)

	(Sister is always relax in every afternoon on the terrace of the house)	
110	*Bola itu **mental** ke arah Budi.* **(The ball is bouncing back towards Budi)**	*Sesuai* (Relevant)
111	*membawa oleh-oleh **tahu** sumedang* **(carrying gift of Sumedang tofu)**	*Sesuai* (Relevant)
112	*Beberapa **kali** anak itu dipukuli oleh ayahnya* **(Several times, the child was beaten by his father)**	*Sesuai* (Relevant)
113	*Daging ayam itu harus dipotong **dadu*** **(The chicken meat must be cut into cubes)**	*Sesuai* (Relevant)
114	***Tahu** itu bulat seperti bola ping pong* **(Tofu is a round-shape as ping pong balls)**	*Sesuai* (Relevant)
115	*sebelum masuk rumah baiknya mengucap **salam*** **(It is better to say greetings before entering house)**	*Sesuai* (Relevant)
116	***pupus** sudah harapanku selama ini* **(my hopes have been broken)**	*Tidak Sesuai* (Irrelevant)
117	*harga daun di **pasar** blimbing cukup murah* **(the price of leaves in the Blimbing market is quite cheap)**	*Sesuai* (Relevant)
118	*pencuri itu **kabur** melarikan diri setelah dipergoki pemilik rumah* **(thieves ran away after being caught by the homeowner)**	*Sesuai* (Relevant)
119	*mata nenek sudah **kabur** karena dimakan usia* **(Grandma eyes are blurred because due to aging)**	(Relevant)
120	*sempat terjadi ke**rusuh**an yang sangat mencekam saat tragedi 1998* **(there was a riot which was very tense during the 1998 tragedy)**	*Sesuai* (Relevant)
121	***adam** adalah manusia laki-laki pertama dimuka bumi* **(Adam is the first male human being on earth)**	*Sesuai* (Relevant)
122	*dia menderita penyakit **asma** semenjak kecil sehingga dia dianjurkan untuk berenang tiap minggu* **(he suffered from asthma since childhood so he was recommended to swim every week)**	*Sesuai* (Relevant)
123	*pak budi sangatlah kaya tapi dia orang yang **kikir*** **(Mr. Budi is very rich, but he is a stingy person)**	*Tidak Sesuai* (Irrelevant)
124	*teman saya mempunyai penyakit **asma** semenjak lama* **(my friends have had asthma since a long time ago)**	*Sesuai* (Relevant)
125	*dia adalah orang yang **kejam** karna memberikan harapan palsu* **(he is a cruel person because he gives false hopes)**	*Tidak Sesuai* (Irrelevant)
126	*setiap hari saya haruslah menyapu **halaman***	*Tidak Sesuai* (Irrelevant)
	(every day I have to sweep the yard)	
127	*warga menjadi **rusuh** akibat haknya tidak terpenuhi.* **(residents became riot because their rights were not fulfilled)**	*Sesuai* (Relevant)
128	*Kelelawar mencari makan di **malam** hari* **(Bats look for food at night)**	*Sesuai* (Relevant)
129	***Rapat** tertutup itu dihadiri oleh jajaran KPK beserta Kapolri* **(The closed meeting was attended by the boards of the Corruption Eradication Commission and the relevant National Police Chief)**	*Sesuai* (Relevant)
130	*Ibu membeli buah **apel** di Pasar* **(Mothers buy apples in the Market)**	*Sesuai* (Relevant)
131	*Ratna menanam **Bunga** Anggrek di dep n rumah* **(Ratna plants Orchid Flowers in front of the house)**	*Sesuai* (Relevant)
132	*dia **gondok** hingga sangat marah karena dibully teman-temannya* **(he got mad until he was very angry because his friends bullied him)**	*Sesuai* (Relevant)
133	*Penyakit **gondok** adalah kondisi dimana terjadi pembengkakan kelenjar tiroid* **(Mumps is a condition in which the thyroid gland swells)**	*Sesuai* (Relevant)
134	***halaman** rumah itu sangat luas sekali* **(That home yard is very spacious)**	*Sesuai* (Relevant)
135	*buka **halaman** 51 dan kerjakan soalnya* **(open page 51 and do the practice)**	*Sesuai* (Relevant)
136	*meski kredit memiliki **bunga** yang tinggi masih saja orang banyak yang memakai* **(Even though credit has high interest, there are people who still use it)**	*Tidak Sesuai* (Irrelevant)
137	***buku** habis gelap terbitlah terang merupakan ciptaan dari R.A Kartini* **(book titled " after the dark rises the brigh is a creation of R. Kartini)**	*Sesuai* (Relevant)
138	*meskipun berbahaya **bisa** ular digunakan untuk membuat penawar racun untuk orang yang digigit.* **(Even though it is dangerous, snake venom can be used to make antidotes for people who are bitten)**	*Sesuai* (Relevant)
139	*dia menjalani **opname** di rumah sakit akibat kecelakaan tempo hari* **(he was hospitalized due to an accident yesterday)**	*Sesuai* (Relevant)
140	*tumbuhan di indonesia memiliki **ragam** yang sangat banyak* **(plants in Indonesia have a very wide variety)**	*Tidak Sesuai* (Irrelevant)

From the 140 tested sentences above, it is found that 30 sentences that contain ambiguous word cannot be defined properly by system. The undefined words are various. For

instance, there are 2 ambiguous words which cannot be defined within a sentence, 1 ambiguous word that is able to be defined and 1 ambiguous word that cannot be defined in sentence.After the test result of 140 sentences containing of 2 ambiguous words is known, it is continued with the accuracy calculation from the system with the succeeding formula:

$$\text{Accuracy} = \frac{\textit{Number of 2 ambiguous words correctly translated}}{\textit{total number of 2 ambiguous words}} \; x\; 100\% \qquad (1)$$

$$\text{Accuracy} = \frac{110}{140}\; x\; 100\%$$

$$\text{Accuracy} = 78.6\%$$

From the accuracy calculation above, it can be concluded that homograph word meaning defining system is successful to define 2 word meaning according to the homograph word in inputted sentences as many as 78.6%.

IV. CONCLUSION

In line with the test which is made in building the homograph word meaning defining system using LESK algorithm, the conclusions which are able to be withdrawn are Homograph data which is obtained from KBBI (*Kamus Besar Bahasa Indonesia*) follows the rule for homograph word marking which is existed at KBBI. The elicited data is in the form of words, meanings, and sentence examples. About 72 homograph words exist and about 150 homograph word meanings also exist. There is an inadequacy in the system when the obtained score is equal, the system cannot determine the homograph word meaning. It is because the LESK algorithm only chooses the highest score to determine the meaning. The system cannot determine the word meaning when a sentence related to the homograph words is not existed, which later will affect to the number and the completeness of the knowledge base.

For further development, it is expected that future researcher to apply more knowledge base and database. It is also expected that the knowledge base and database is more valid. Thus, it will make the system easier to determine word meaning if there is valid resource available. For the rule of creation, the score is equal when determining the word meaning. This should be existed in order to make the system determine the meaning if d equal score is found. Mutual Post tagging addition will become the increment in system accuracy.

ACKNOWLEDGMENT

This research is partially supported by Laboratorium Informatika Universitas Muhammadiyah Malang. Authors wish to thank Universitas Muhammadiyah Malang for providing the funding.

REFERENCES

[1] S. Lamb, "Lexicology and Semantics," in *Voice of America Forum Lectures*, 1969.

[2] G. Keraf, *Diction and Language Style, I^st edition*. Nusa Indah, 1981.

[3] P. Wijana, "Reasons for the Establishment of Homonyms in Indonesian," *Univ. Gajah Mada*, 2013.

[4] P. Chen, W. Ding, C. Bowes, and D. Brown, "A Fully Unsupervised Word Sense Disambiguation Method Using Dependency Knowledge," *Comput. Linguist.*, no. June, pp. 28–36, 2009.

[5] L. Tan, "Examining Crosslingual Word Sense Disambiguation," 2013.

[6] J. Harmoejanto *et al.*, "Formation of Graphs for Labeling Meanings of Words Based on Synset, Synset and Gloss Relations," pp. 0–5, 2010.

[7] S. Banerjee, "Adapting the LESK Algorithm for Word Sense Disambiguation to WordNet," no. December, 2002.

[8] Pateda, *Lecsical Semantic*. Jakarta: Rineka Cipta, 2001.

[9] J. Arukgoda, V. Bandara, S. Bashani, V. Gamage, and D. Wimalasuriya, "A Word Sense Disambiguation Technique for Sinhala," in *2014 4th International Conference on Artificial Intelligence with Applications in Engineering and Technology*, 2014, pp. 207-211.

[10] R. Sawhney and A. Kaur, "A modified technique for Word Sense Disambiguation using Lesk algorithm in Hindi language," in *2014 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, 2014, pp. 2745-2749.

[11] M. Lesk. Automatic sense disambiguation using machine readable dictionaries: How to tell a pinecone from a ice cream cone. In Proceedings of SIGDOC '86, 1986.

[12] D. D. Purwanto, "Synonym and Word Sense Disambiguation to Complete the Detector of Plagiarism for Final Project Documents," *J. Inf. Syst.*, vol. 11, no. 1, pp. 33–38, 2015.

[13] R. M. Gitasari, "Removal of Ambiguity in the Meaning of Words in Indonesian Sentences by Using a Parser , Wordnet Dan Algoritma Lesk Word Sense Disambiguation in Indonesian Sentence Use Parser , Wordnet and LESK algorithm," 2007.

[14] , D. R., & Rosyadi, A. R.. Paragraph Selection Methods Using Feature-Based On Segment-Based Clustering Process Using Paragraphs For Identifying Topics Indications Detection of Plagiarism System. *KINETIK*, *3*(2), 91-100, 2018.

[15] Basuki, S., Rizky, A., & Wicaksono, G. W. Case Based Reasioning (CBR) for Medical Question Answering System. *Kinetik: Game Technology, Information System, Computer Network, Computing, Electronics, and Control*, *3*(2), 113-118, 2018.

WSD

ORIGINALITY REPORT

6%	5%	4%	1%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Setio Basuki, Yufis Azhar, Agus Eko Minarno, Christian Sri Kusuma Aditya, Fauzi Dwi Setiawan Sumadi, Ardiansah Ilham Ramadhan. "Detection of Reference Topics and Suggestions using Latent Dirichlet Allocation (LDA)", 2019 12th International Conference on Information & Communication Technology and System (ICTS), 2019 Publication	2%
2	id.wiktionary.org Internet Source	2%
3	kbbi.kata.web.id Internet Source	<1%
4	Hardianto Wibowo, Dimas Nurpratama, Wildan Suharso, Agus Eko Minarno, Galih Wasis Wicaksono, Dani Harmanto. "Impact Evaluation of Procedurally Content Generated Against Immersion Games Using ANOVA", 2020 8th International Conference on Information and Communication Technology (ICoICT), 2020 Publication	<1%

5	zh.scientific.net Internet Source	<1%
6	desidwiingkana.blogspot.com Internet Source	<1%
7	www.proceedings.com Internet Source	<1%
8	insightsociety.org Internet Source	<1%
9	thesai.org Internet Source	<1%
10	lbeifits.files.wordpress.com Internet Source	<1%
11	www.ijcaonline.org Internet Source	<1%
12	kinetik.umm.ac.id Internet Source	<1%
13	www.ed.state.nh.us Internet Source	<1%

Exclude quotes	Off	Exclude matches	Off
Exclude bibliography	Off		