The Effectiveness of Two-Phase Translation Method compared to Every-Match Method in Vocabulary Translation

Matahari
mataharitarie@gmail.com

Universitas Pendidikan Muhammadiyah Sorong

ABSTRACT

The use of a dictionary will be more efficient with a digital dictionary because it may be used anytime and wherever we are. Vocabulary search in digital dictionaries can use several search methods or vocabulary translators, including the Every-Match translator method and the Two-Phase translator method. The Every-Match and Two-Phase translator methods were used to search on Indonesian, English, and Arabic vocabulary in this research. The accuracy of the two methods will then be analyzed to see which is the more accurate translation method. The experimental methodology is used in this study, with the aim to determine the effect of a treatment on the experimental group's results. The Every-Match and Two-Phase translation methods are used in this multilingual digital dictionary. A process of matching keywords with a database in this digital dictionary system. Keywords in Indonesian will be used as a basis for searching in this digital dictionary. The search will be divided into two phases: a search using the same keywords and a search using databases that have keywords that are similar to the ones entered. The goal is to avoid keyword typing errors. And, based on the results of the data analysis, the Mann Whitney U test shows that there is a very significant difference between the Every-Match method and the Two-Phase method in translating words from Indonesian in English to Arabic. The average value of Every-Match was 34.8 or 70%, and the translation using the Two-Phase method was 41.7 or 83%.

Key words: Dictionary, Every-Match, Two-Phase, Method, Translation.

INTRODUCTION

Language is one of the communication media used by humans to interact with each other, and language cannot be separated from human life. One way to learn a language is to study the vocabulary itself, by knowing and understanding a person's vocabulary it can be easier to learn the language, and a dictionary is one of the tools in learning a language. A dictionary is a reference book that contains words and expressions, usually arranged in alphabetical order with information about their meaning, usage, or translation.

With today's advances in technology, dictionaries are not only in printed form, but are more developed towards electric or digital dictionaries. With the digital

dictionary, the use of the dictionary will be more efficient because it can be used whenever and wherever we are. The stages of the translation process in the digital dictionary refer to the Machine-readable Dictionary which is used to translate vocabulary by translating word by word, and using the Every-Match method. The Every-Match method can produce multiple translations, and result in ambiguous translation results. The Two-Phase translator method was established to solve these issues. The Every-Match and Two-Phase translator methods will be used to search Indonesian, English, and Arabic vocabulary in this research's translator method in the digital dictionary. The accuracy of the two methods will then be compared to determine which is the more accurate of the two translator methods.

Based on the thoughts that have been described, an objectives is obtained, that is optimizing the search for each word searched in a multi-language Indonesian-Arabic-English digital dictionary and displaying the results, is there any difference in the vocabulary search results between using the Each-Match method and the Two-Phase method. And between these two methods, which method is the most accurate.

The purpose of this research was to determine the difference in vocabulary search results between the use of the Every-Match method and the Two-Phase method, and to determine the most accurate method between the two methods in translating vocabulary from Indonesian to English and Arabic.

LITERATURE REVIEW

Previous Related Study

The previous related study must contain at least three previous studies related with the issue discussed.

There are several previous related studies focusing on translation methods in digital translation, such as in the research conducted by Richard Frankel, Jared Jennings, and Joshua Lee (2021), they compare the ability of dictionary-based and machine-learning methods to capture disclosure sentiment. It was found that measures based on machine learning offer a significant improvement in explanatory power more than dictionary-based measures. Overall, their results suggest that machine-learning methods offer an easily implementable, more powerful, and reliable measure of disclosure sentiment than dictionary-based methods. The other research conducted by Ivan Dunđer (2020), in this research focused on experimented with phrase-based statistical machine translation for English-Croatian pairs. Researchers focusing on machine translation methods from Croatian have great potential. Further investigation should yield definite conclusions about what domain adaptation methods are suitable for translating industry-specific texts, from English to Croatian, and vice versa. And research conducted by Tatiana Lenskaia (2021) which focused on dictionary-based methods. This research indicates that dictionary-based methods are a powerful computational approach for analyzing unannotated genome data. The methods are scalable and effective for large-scale screening research. This allows researchers to explore and analyze large amounts of genome data that are constantly added to the databases.

Linguistic

Linguistics is the science of language or the study of language, which comes from the Latin word lingua which means language. Linguistics does not study only one language but related to language in general term. (Suyudi Ichwan. 1997). Word itself has two basic meanings that language as a general concept, and a language itself.

Another definition of language is as a communication system that allows humans to work together. This definition emphasizes the social function of language and the fact that humans use it to express themselves and to manipulate objects in their environment. Language can refer to a special human capacity to acquire and use a complex communication system, or to a specific instance of a complex communication system.

Lexicography

Lexicography is a branch of the field of linguistics that examines and studies the way/technique of making/compiling a dictionary, which consists of designing, compiling, using and evaluating the dictionary itself. The term lexicon in linguistics means a vocabulary of words that are often called lexemes. Every language has a fairly large vocabulary, covering tens of thousands of words and even more, and each word has its own meaning or meaning.

In history, there are several big names who started the compilation of dictionaries, namely Samuel Johnson (1709-1784) and Noah Webster (1758-1843). In 1755, Samuel Johnson, who was an English linguist, created a two-volume Dictionary of The English Language. And Noah Webster, in America in 1828, made the first two volumes of An American Dictionary of the English Language. After that, it was published by the Oxford English Dictionary consisting of 12 volumes in 1884.

Translator Method

The Every-Match method is the method used by the Machine-readable Dictionary (MRD). The translation process can produce a word that may have more than one meaning. By using this method, all possible translation results will be displayed, which causes the result of the translation process to experience a high level of ambiguity. Two-Phase Method is a method by going through two stages in its implementation. By using Two Phase, the translation is carried out in two directions, namely the translation from the source language to the target language, then reversed from the target language to the source language. The results of the translation used are those that produce words according to the source language that are translated back.

The Two-Phase method does not use all the translation results, but only a few translation results. The basic assumption of this method is that the translation of the results of the translation of a word must produce the original word or its input. If this happens, then the translation results are considered valid. The first stage is

looking for input words and calculating the level of similarity by taking the root word (from Indonesian vocabulary) and then looking for a translation, as in the following example:

Input : makan
Search keywords : memakan, makanan, makan, dimakan
Keyword found : makan
Translate into Arabic : أكل

After the initial stage is complete, namely finding the word you are looking for and its translation, then proceed to the next stage. That is looking for an English translation, because at first the keywords have been filtered and a suitable word is found, then this second stage is only looking for an English translation, namely: makan = eat. These two stages will make it easier to search if the input word is reversed, such as English-Arabic-Indonesian or Arabic-Indonesian-English with reference to Indonesian input as the key word.

METHOD

Design and Samples

The experimental method is applied in this research, which is a method that includes manipulating the research object and the existence of a control condition (Nazir, 1983). In other words, this method aims to determine the effect of a treatment on the results of the experimental group. This research was conducted at the Universitas Pendidikan Muhammadiyah (UNIMUDA) Sorong in May 2021. The UNIMUDA Sorong was chosen as the research location as it is a representative university in West Papua's province.

The methods used in multi-language digital dictionaries are the Every-Match method and the Two-Phase method. The system of this digital dictionary is a process of matching keywords to the database. The input to search for a word in this digital dictionary is the entry of keywords in the form of Indonesian, which will be used as the basis for searching. The search will go through two stages, namely a search based on the same keywords and a search on a database that has similarities to the keywords entered. The aim is to avoid miswriting the keywords. The words obtained from the search results in the first stage (taken from the keyword entry results) will be displayed, and the results from the second stage (words that are similar to keywords) will be calculated for their similarity level which will be used as a filter.

If the similarity level is above 70%, then the word passes the filtering which will then be displayed along with its translation in Arabic or English. The process of translation has two stages (two-phase), namely the input of words that were originally Indonesian and the translation into Arabic. In this search, the level of similarity of the keywords entered is found. After the next stage of Arabic is found, the word (Indonesian translation) which refers to English. This will make it easier

if the word is reversed from Indonesian-Arabic-English or from English-Arabic-Indonesian. It will also make it easier to search. The word you are looking for will immediately find its Arabic and English translation. The flowchart and stages of the Two-Phase method can be seen in Figure 1.

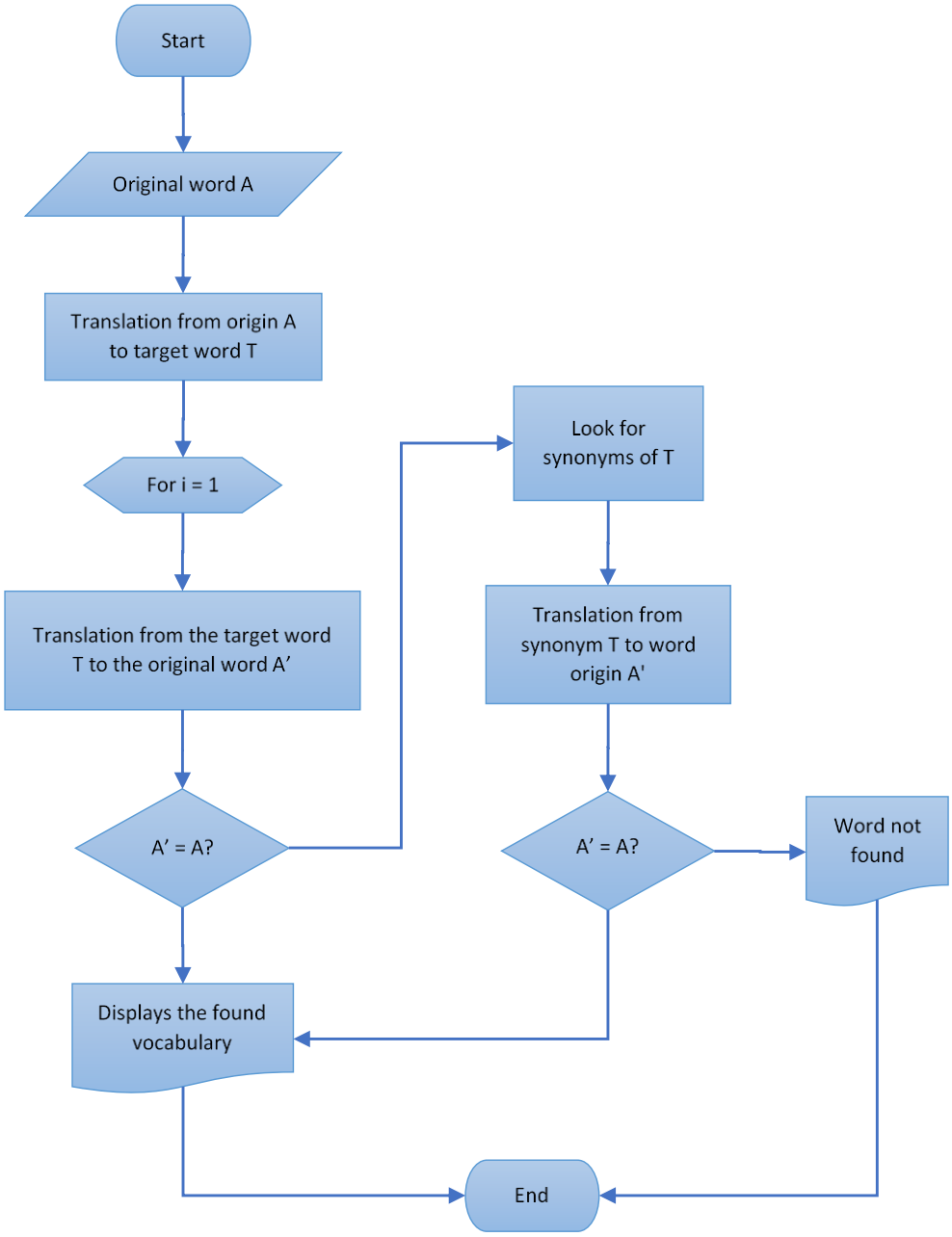

Figure 1. Two-Phase Translator System Design

Flowchart in Figure 1 shows the flow of translation using Two-Phase Method. With a notation that represents each process can be explained as follows:

1. Variable A is used to represent the word to be translated, or also called the original/input word.

2. Variable T is the result of translation in the target language from the original word A.
3. Variable i = 1, is a parameter for true value if T is found.
4. Variable A' is the result of the translation of the target word T, which is reversed by the translation path.
5. Variable A" is the result of the translation of the synonym T

The Two-Phase method in Figure 1 can be implemented as follows:
1. When the origin word A is submitted, it will be translated into the target language T.
2. When the translation result of T is found, the translation process will be reversed with T as the initial language and A as the target language, and will produce A'.
3. If the result of translation A' is the same as the original language of A, then the translation is considered valid, and the result T will be displayed so that the process will be completed.
4. And if the result of the translation of A' is not the same as the original language of A, then we will look for synonyms of T that are close to the original language of A.
5. When the result is found A", the word will be displayed. And if not, a message will appear the word not found, and the process is complete.

In some cases there can be a situation where none of the translation candidates is found. in that case it can be modified as follows:
1. If the English word E does not produce the same word as Arabic A, then: find the synonym of the English word, then translate it into Arabic using the every-match method, each translation of the synonym produces a word that matches the word origin A is a candidate for translation
2. If there is no English word E or its synonym that produces the original word A, then use the first match method of E as a translation candidate

Instrument and Procedure

The data collection was carried out based on the experimental design, which was determined by two groups of respondents consisting of students of the Program Studi Pendidikan Teknologi Informasi (PTI) UNIMUDA Sorong who had relatively the same characteristics, each totaling 10 respondents. What is meant by the same characteristics is that a respondent in both experimental groups is in the 6th semester with an average achievement index of 3.00-3.30.
The first experimental group consisting of 10 students was given the task of finding the translation of fifty Indonesian words into English and Arabic using the Every-Match method. While the second experimental group consisting of 10 students was given the task of finding the translation of 50 Indonesian words into English and Arabic using the Two-Phase method. The time given to both groups to find the translation of fifty vocabulary words is 1 minute. Within 1 minute, we want to know

how many correct translation words were found by the two experimental groups. The sample of fifty Indonesian vocabularies that are sought for translation is listed in Table 1.

Table 1. Samples of Indonesian vocabulary and English and Arabic translations (sample contains 25 vocabulary)

No	Indonesian Vocabulary	Translation	
		English	Arabic
1	Absen	Absent	غائب
2	Ada	Exist	وجود
3	Adalah	There Is	يكون
4	Adanya	Availability	التوفر
5	Adapun	As For	أما بالنسبة لل
6	Agung	Great	عظيم
7	Air	Water	ماء
8	Ajang	Event	حدث
9	Akan	Will	وسوف
10	Akhir	End	نهاية المطاف
11	Akhirnya	Finally	في نهاية المطاف
12	Akibat	Result	نتيجة لذلك
13	Aksara	Script	النصي
14	Aksarawan	Librarian	أمين المكتبة
15	Aksi	Action	الإجراء
16	Alam	Nature	طبيعة سجية
17	Alami	Experience	خبرة
18	Alamat	Address	عنوان
19	Alas	Base	يتمركز
20	Alasan	Reason	من الأسباب
21	Alat	Tool	الجهاز
22	Ambang	Threshold	عتبة
23	Ambil	Take	يأخذ
24	Ampas	Dregs	الثمالة
25	Amplop	Envelope	مغلف

Then the average vocabulary search was generated from the two experimental groups, namely the Experimental Group I (using the Every-Match Method) and the Experimental Group II (using the Two-Phase Method). From the average results, the Mann-Whitney U Test will then be analyzed and performed to see the differences produced by the two methods, and to find out which of the two methods is the most accurate.

Data Analysis

To find out the difference in translation results between the Every-Match method and the Two-Phase method, the Mann-Whitney U Test was carried out, which is a test to see the difference in the mean of two data groups. These two data groups are the data generated by the Experimental group I using the Every-Match Method and the Experimental group II using the Two-Phase Method, and each group consists of 10 students of Program Studi Pendidikan Teknologi Informasi (PTI) Universitas Pendidikan Muhammadiyah (UNIMUDA) Sorong. Meanwhile, to find out which method is better in doing the translation, the mean of the correct and correct translation results from the two experimental groups is calculated and the highest value is chosen.
The Mann-Whitney U Test is used to determine the difference in translation outcomes between the Every-Match and Two-Phase methods. It begins with the creation of a worksheet for ranking and calculating $\sum R_1$, $\sum R_2$, U_1, and U_2. The values $\sum R_1 = 55$ and $\sum R_2 = 155$ will be generated by the worksheet. The following information and calculations will be obtained as a result:

Is known:

Number of respondents in Experimental Group I (Every-Match Method):
$n_1 = 10$

Number of respondents in Experimental Group II (Two-Phase Method):
$n_2 = 10$

$n = n_1 + n_2 = 20$

Level Significaance:
$\alpha = 0{,}05$

$\sum R_1 = 55$
$\sum R_2 = 155$

Calculate U:

$$U_1 = n_1 n_2 + \frac{n_2(n_2+1)}{2} - \sum R_2$$
$$= (10)(10) + \frac{10(10+1)}{2} - 155$$
$$= 0$$

$$U_2 = n_1 n_2 + \frac{n_1(n_1+1)}{2} - \sum R_1$$
$$= (10)(10) + \frac{10(10+1)}{2} - 55$$
$$= 100$$

RESULT AND DISUSSION

Then the value of U is calculated, resulting in the value of $U_1 = 0$, and $U_2 = 100$. The calculated U is determined by selecting the smallest U value, namely $U_1 = 0$, the calculated U value $= 0$. The calculated U value will be compared with the value in the Mann Whitney U table (see Table 2) by testing two data at the $\alpha = 0.05$ level and referring to the number of variables m and variable n, namely U at $\alpha = 0.05$ with a value of m=10 and a value of n=10 resulting in a value of 23. Because U count $< U$ table, it is concluded that there is a very significant difference between the Every-match method and the Two-Phase method in the translation of words from Indonesian to English.

Table 2. Table U Mann Whitney, Test 1 variable at Level 0.025 and Test 2 variables at Level 0.05

m	n																			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	—																			
2	—	—																		
3	—	—	—																	
4	—	—	—	0																
5	—	—	0	1	2															
6	—	—	1	2	3	5														
7	—	—	1	3	5	6	8													
8	—	0	2	4	6	8	10	13												
9	—	0	2	4	7	10	12	15	17											
10	—	0	3	5	8	11	14	17	20	23										
11	—	0	3	6	9	13	16	29	23	26	30									
12	—	1	4	7	11	14	18	22	26	28	33	37								
13	—	1	4	8	12	16	20	24	28	33	37	41	45							
14	—	1	5	9	13	17	22	26	31	36	40	45	50	56						
15	—	1	5	10	14	19	24	29	34	38	44	50	54	60	54					
16	—	1	6	11	15	21	26	31	37	42	47	53	59	64	70	75				
17	—	2	6	11	17	22	28	34	39	45	51	57	63	69	75	81	87			
18	—	2	7	12	18	24	30	36	42	48	55	61	67	74	80	86	93	99		
19	—	2	7	13	19	25	32	38	46	52	58	65	72	78	85	92	99	100	113	
20	—	2	8	14	20	27	34	41	48	55	62	69	76	83	90	98	105	112	119	127
21	—	3	8	15	22	28	36	43	50	58	65	73	80	88	96	103	111	119	126	134
22	—	3	9	18	23	30	38	45	53	61	69	77	85	93	101	109	117	125	133	141
23	—	3	9	17	24	32	40	48	56	64	73	81	89	98	106	116	123	132	140	149
24	—	3	10	17	25	33	42	50	59	67	76	85	94	102	111	120	129	138	147	156
25	—	3	10	18	27	35	44	53	62	71	80	89	98	107	117	126	135	145	154	163
26	—	4	11	18	28	37	48	55	64	74	83	93	102	112	122	132	141	151	161	171

27	—	4	11	20	29	38	48	57	67	77	87	97	107	117	127	137	147	158	168	178
28	—	4	12	21	30	40	50	60	70	80	90	101	111	122	132	143	154	164	175	186
29	—	4	13	22	32	42	52	62	73	83	94	105	116	127	138	149	160	171	182	193
30	—	5	13	23	33	43	54	65	76	87	98	109	120	131	143	154	166	177	189	200
31	—	5	14	24	34	45	56	67	78	90	101	113	125	136	148	160	172	184	196	208
32	—	5	14	24	35	46	58	69	81	93	105	117	129	141	153	166	178	190	203	215
33	—	5	15	25	37	48	60	72	84	98	108	121	133	146	159	171	181	197	210	222
34	—	5	15	28	38	50	62	74	87	99	112	125	138	151	164	177	190	203	217	230
35	—	6	16	27	38	51	64	77	89	103	116	129	142	158	169	183	196	210	224	237
36	—	8	18	28	40	53	66	79	92	106	119	132	147	161	174	188	202	218	231	245
37	—	8	17	29	41	55	68	81	95	109	123	137	151	165	180	194	209	223	238	252
38	—	8	17	36	43	58	70	84	98	112	127	141	156	170	183	200	215	230	245	250
39	0	7	18	36	44	58	72	86	101	115	130	145	160	175	190	206	221	236	252	257
40	0	7	18	31	45	60	74	89	103	119	134	149	165	180	198	211	227	243	258	274

Because U count < U table, it is concluded that there is a very significant difference between the Every-Match method and the Two-Phase method in word translation from Indonesian-English-Arabic. The translation results of the Experimental Group I using the Every-Match method obtained an average value of 34.8 or 70%, and the translation results of the Experimental Group II using the Two-Phase method were 41.7 or 83%. From these results, it can be determined that the Two-Phase method has a better translation accuracy than the Every-Match translator method.

CONCLUSION

By using the Two-Phase method, the search and translation process is more accurate and effective than the Every-Match method, because the Two-Phase method uses two stages in one system. First, using the Every-Match translator method which checks every database record after that in order to get a high level of accuracy, the next stage of the search is used by using filters to match the search results with keyword input. The search stage is not carried out only on one language data, but is carried out on two other data, namely English and Arabic with Indonesian as the primary key or index. Word indexing is done in alphabetical order to facilitate search and translation.

And from the results of the data analysis that has been carried out, it can be concluded that there is a very significant difference between the Every-Match method and the Two-Phase method in word translation from Indonesian-English-Arabic resulting from the Mann Whitney U Test. Then, the translation results of the Experimental Group I using the Every-Match method obtained an average value of 34.8 or 70%, and the translation results of the Experimental Group II using the Two-Phase method were 41.7 or 83%. From these results, it can be determined that the Two-Phase translator method is convincingly better at the level of translation

accuracy than the Every-Match translator method and can improve the accuracy of the translation results in the Indonesian-English-Arabic digital dictionary.

REFERENCES

Aljlayl, Mohammed. & Frieder, Ophir. (2001). *Effective Arabic-English Cross-Language Information Retrieval Via Machine-Readable Dictionaries and Machine Translation: Proceedings of the Tenth International Conference on Information and Knowledge Management, Atlanta Georgia, 5 – 10 October 2001* (pp. 295-302). New York, United States: Association for Computing Machinery

Djamaries, J ST. (2012). *Kamus Besar Bahasa Inggris (Inggris-Indonesia, Indonesia-Inggris) Cetakan Keenam*. Jakarta, Indonesia: Citra Harta Prima.

Dunđer, Ivan. (2020). Machine Translation System for the Industry Domain and Croatian Language. *Journal of Information and Organizational Sciences*, 44(1), 2020, 33-50. doi: https://doi.org/10.31341/jios.44.1.2

Frankel, Richard M. Jennings, Jared N., & Lee, Joshua A. (2021). Disclosure Sentiment: Machine Learning vs Dictionary Methods. *Management Science*, 67(11), 1526-5501. doi: https://doi.org/10.1287/mnsc.2021.4156

Hurskainen, A. (2009). *Two-Phase Implementation of Morphological Analysis*. Technical Reports in Language Technology. Report No 6, 2009. Retrieved from http://www.njas.helsinki.fi/salama/technical-reports.html

Kaji, Hiroyuki., Kida, Yuuko., & Morimoto, Yasutsugu. (1992). *Learning Translation Templates from Bilingual Text*: *Proceedings of the 14th International Conference on Computational Linguistics (COLING-92), Nantes, 23−28 August 1992* (pp. 672−678). Stroudsburg, United States: Association for Computational Linguistics.

Kim, Sungwoo. (2019). Playing with Machine Translation in Language Classroom: Affordances and Constraints. *Multimedia-Assisted Language Learning*, 22(2), 9-28. 1229-8107. doi: 10.15702/mall.2019.22.2.9.

Kridalaksana, Harimukti. (2001). *Pembentukan Kata dalam Bahasa Indonesia*. Jakarta, Indonesia: Gramedia Pustaka Utama.

Lenskaia, Tatiana. (2021). *Dictionary-Based Methods and Their Applications in Biology and Medicine*. (ProQuest Dissertations). University of Minnesota, Minnesota, USA.

M, Matahari. (2013). *Pembuatan Software Kamus Multi Bahasa Indonesia-Arab-Inggris dengan Menggunakan Metode Penerjemah Two Phase*.

(Undergraduate thesis), Universitas Islam Negeri Maulana Malik Ibrahim Malang, Malang, Indonesia.

Nazir, Moh. (2014). *Metode Penelitian Cetakan Kesepuluh*. Jakarta, Indonesia: Ghalia Indonesia.

Sánchez, Noé Alejandro Castro. (2020). Processing of Semantic Relationships in a Monolingual Dictionary for Creation of Machine-Readable Resources. *Computación y Sistemas*, 24(4), 1571-1580. doi: 10.13053/CyS-24-4-3876.

Sudjana. (2005). *Metoda Statistika Cetakan Keenam*. Bandung, Indonesia: Taristo.

Suyudi, Ichwan.(1977). *Pengantar Linguistik Umum*. Depok, Indonesia: Penerbit Gunadarma.

Utomo, Adi Heru., & Setyohadi, Dwi P. Sarwo. (2014). Implementasi Two-Phase Translation Method Pada Pembuatan Web Program Transitive Translation Antara Bahasa Jawa Dan Bahasa Inggris Menggunakan Pivot Bahasa Indonesia. *Jurnal Teknologi Informasi dan Terapan*, 01(01), 2580-2291. Retrieved from http://jtit.polije.ac.id/index.php/jtit/article/view/3