REPRESENTATION OF INDONESIA IN WIKIPEDIA

Sunarsih

Institut Teknologi Sumatera

sunarsih@dkv.itera.ac.id

ABSTRACT

The objective of this study is to describe the representation of Indonesia in a Wikipedia article. The approach applied in this study is corpus-based discourse analysis. This approach concerned with the keyword "Indonesia" and how it conveyed meaning related to the other words around (collocation) in the left and right side of the word "Indonesia". By analyzing the collocation, the representation of Indonesia is interpreted and input as the meaning property of the word "Indonesia". The finding of this study shows that the word "Indonesia" in Wikipedia article contained meaning components such as Indonesia is a country whose rich cultural and natural resources, long history; developing technology; developing economy; facing some problem; actively participating in many international affairs and sports competition. It implies that, in general, the image of Indonesia in Wikipedia is represented positively.

Keywords: Indonesia, Wikipedia, corpus-based discourse analysis

I. Introduction

Wikipedia is the most popular and free online encyclopedia used today by people all over the world. People can read, access, write, and edit articles collaboratively and voluntarily in Wikipedia. At the moment this study is being written, Alexa ranks Wikipedia in the fourteenth position of the most visited site. There are more than 54 million articles, written by more than 130 thousand contributors from more than 300 of various language users. This openness towards new users was cited as both a source of strength and a ource of weakness; however, Wikipedia has grown exponentially in users and information since 2002 and has been highlighted as a success story of shared knowledge systems at low cost (Kittur et.al, 2006). Aside from its reputation on the accuracy of the published articles, it becomes a well-known habit that if we want to know something in general, Wikipedia is the first door to open and supply the knowledge. We could find many articles about countries around the world, including the one about Indonesia.

If we type the word "Indonesia" in Google search bar, the top result is suggesting us to visit Wikipedia. The content of the article includes the etymology, history, geography,

government and politics, economy, demography, and culture. This paper observes the representation of Indonesia in the article written in Wikipedia entitled "Indonesia". It is essential to know how people might have collectively common knowledge about Indonesia from the most accessible encyclopedia in the world. Since Wikipedia's articles form a highly interlinked knowledge base, they might contains a rich body of semantic nuance that involves knowledge of persons, places, specific words or word-senses (Zesch et. al, 2008). This study describes the representation of Indonesia as written in Wikipedia. What has been written there might construct a certain image about Indonesia in the mind of common people, based on the word-senses of the word collocation around the keyword being analyzed. The research is carried out under the corpus-based discourse analysis method.

This blends corpus linguistic approaches with discourse analysis. Corpus is an electronically generated set of texts (Gabrielatos, 2014:3). Corpus processing is a computational procedure that processes linguistic data to find linguistic patterns that enable researchers to deduce how language is used in the construction of discourse (Baker, 2008: 1). Corpus linguistics is the study of language as appearing in the corpus of electronic texts. It is to observe linguistic patterns as provided in the electronic texts, such as collocation, keywords, frequency, word groups, word distribution, and so on. The approach used in this research is to use the keyword analysis by concordance and collocation.

The keyword, in this study, refers to the particular vocabulary which is selected to be the research query. In this case, the keyword is "Indonesia" in the article about Indonesia in Wikipedia. The list of a word or phrase, with a few words of context either side of it, so that we can see at a glance how the word tends to be used is called as concordance (Baker, 2010). In other words, concordance is a list of examples of keyword usage in the article being analyzed or the list of the keyword in context (KWIC). In a concordance, the sequence of words is not merely a complete sentence. The series of words in a concordance is not merely a full sentence. The word limit to the right and left of the keyword to be analyzed is predetermined by the researcher's input to the corpus software program, *AntConc*. After receiving the input, the concordance feature will display the overall presence of a certain keyword in a corpus, displayed in the context or sense in which the word is used, i.e, along with several other words to the right and left of the keyword being studied. Generally, there are 4-5 words to the right and left of the keywords to be analyzed. Words other than within that spectrum may be invisible. A concordance of a keyword can show and express meaning, opinion, and certain linguistic patterns. How the meaning of a keyword in a text is constructed can be scrutinized through the concordance (Al Fajri, 2017). The concordance also refers to keywords in context. The keywords referred to in this case are terms that occur frequently or relevant terms that are important to be studied by the researcher.

Collocation is the presence of words along with the keyword being studied. Baker (2008: 96) argues that collocations are words that come together with a keyword. The presence of the collocation is usually in one or the closest sentence with the keywords to be analyzed. Collocations can generally range from four to five words on the right and left side of the keywords being analyzed. Hunston (2002: 109) adds that collocation can form an association pattern based on a large number of texts which human intuition or consciousness cannot attain. These patterns are semantic and may contain implicit messages. This potential is usually used by the computer programmer to represent the meaning of a text in higher context, as conducted by Gabrilovich and Markovitch as they derived the corpus from Wikipedia (2007). Collocation can show the core elements of meaning, usage, or association of a word. Collocation is a way to understand the meaning and association between words. This is in line with the idea that the word conveys meaning only if it is present in a context. Therefore, in order to understand the meaning of a word, we must link and compare it in word-to-word relations. Collocation is useful for summarizing the relationships between words in a corpus. Collocation is also useful for finding mainstream discourse.

Researches on corpus-based discourse analysis usually discusses the power relation through the representation in a text. Al Fajri (2017) focuses on how immigrants are represented in hegemonic discourse. Sunarsih (2017) studies the representation of Israel power in Britannica. This study is to describe the meaning or discourse which might be constructed by the concordance and collocation of the keyword "Indonesia" in Wikipedia article entitled *Indonesia*. The word "Indonesia" is mentioned there as many as 174 times. This study does not tend to observe how powerful or how hegemonic Indonesia is. It is to find out the image of Indonesia as represented in Wikipedia article.

II. Methods

This study is a qualitative research. The method applied is corpus-based discourse analysis.   The data in the form of written text is collected from Wikipedia article entitled *Indonesia*. The article is input into a corpus reader software called *AntConc*. The writer put the keyword "Indonesia" as the entry point to analyze the data. The software displayed the frequency of the word "Indonesia" used in the article. Besides, the software showed collocation or the other words on the right and left side of the keyword "Indonesia", which might deliver and contribute a meaning nuance to the keyword "Indonesia".

The data analysis is conducted by reading each lines in the concordance one by one. Then, the lines are categorized based on the most dominant discourse to the least dominant one. It is determined by the number of the concordance list. The concordance with similar notion of discourse or meaning will be put into similar group. The writer interprets the data to discover the meaning components of the word Indonesia which is generated from the interaction between the

word "Indonesia" and the adjective or noun phrases in its left and right sides (collocation). The meaning components contained in the word "Indonesia" become the representation of "Indonesia" in Wikipedia which might construct a certain image in the mind of the readers of Wikipedia. The data is presented in the display of concordance table and the description will follow below the table.

III. Findings and Discussion

After inputting the article *Indonesia*, *AntConc* displayed a concordance or list of how the keyword "Indonesia" used in the Wikipedia article. There found the keyword "Indonesia" used as many as 174 times. The collocation or the words used around the keyword "Indonesia" was studied to find out the meaning which might contribute.

Based on the interpretation conducted, there are eight major patterns of meaning components contained by the word "Indonesia", as the result of the word Indonesia and its collocation. Components of meaning contained by the word "Indonesia" are (1) culturally rich; (2) naturally rich; (3) active participation in international sports competition; (4) developing economy; (5) long history; (6) facing some problems; (7) developing technology; (8) active participation in many international affairs; and. The description of each pattern of meaning components is elaborated below.

Indonesia as a Culturally Rich Country

This meaning component is constructed from 19 concordances of 174. It is formed by the contribution of the words around the keyword "Indonesia". The words around the keywords that form the image as a country full of cultural wealth can be seen via this table 1.

Based on the data of concordance in Table 1, it is listed that the image of Indonesia as a country rich of culture constructed from the collocation or the words and adjective or noun phrases that occur around the keyword "Indonesia", such as *dance, distinct native ethnic, original indigenous cultures, multicultural, multilingual, multi-ethnic, dynamic tradition, various style of clothing, cultural tourism, UNESCO World Heritage Sites, traditional sports, history of tribal, ethnically diverse country, architecture, arts, music, national costume, languages, Hinduism, ethnic diversity, culture of Indonesia, cultural history, Hindi influence, music genres, Malay orchestras, ethnic groups, sculptures,* and *tribal art*.

Table 1. Concordance of Indonesia as a Culturally Rich Country

No	Concordance
1	:Music of Indonesia, Dance in **Indonesia**, and National costume of Indonesia
2	between the two. **Indonesia** consists of hundreds of distinct native ethnic
3	original indigenous cultures. **Indonesia** currently holds ten items of UNESCO'
4	As a result, modern-day **Indonesia** has a multicultural, multilingual and multi-ethnic
5	to be a living and dynamic tradition. **Indonesia** has various styles of clothing as a result
6	destinations for cultural tourism.[195] **Indonesia** has nine UNESCO World Heritage Sites,
7	examples of traditional sports in **Indonesia**. In areas with a history of tribal
8	Indonesians, and Languages of **Indonesia** Indonesia is an ethnically diverse country
9	Indonesian art and Architecture of **Indonesia** Indonesian arts include both age-old art
10	in Indonesia, and National costume of **Indonesia** Indonesian music and dance. Clockwise
11	Languages of Indonesia **Indonesia** is an ethnically diverse country, with around 300
12	6th century,[232] and its history in **Indonesia** is closely related to that of Hinduism
13	with a rich cultural heritage reflecting **Indonesia**'s dynamic history and ethnic diversity
14	Culture Main article:Culture of **Indonesia** See also:Public holidays in Indonesia
15	Public holidays in **Indonesia** The cultural history of the Indonesian archipelago
16	popular music genres in **Indonesia** with notable Hindi influence as well Malay orchestras
17	Main articles: Music of **Indonesia**, Dance in Indonesia, and National costume
18	Main articles: Ethnic groups in **Indonesia**, Native Indonesians, and Languages of
19	megalithic sculptures in **Indonesia**.[273] Subsequently, tribal art has flourished

The readers of article "Indonesia" in Wikipedia might conclude that Indonesia is wealthy of culture, whether it is the tangible or the intangible form. Indonesia is represented as a country possessing various tribes with its own cultural properties such as dance, music, costume, language, sport, history, architecture, faith influence, and arts. This construction tends to contribute a positive image of Indonesia in the readers of Wikipedia or common people who want to know more about Indonesia by accessing the information from Wikipedia. It might be a useful and indirect promotion on the good side of Indonesia. Although it has diverse tribes and beliefs, Indonesia population can live in a harmony and peace.

Indonesia as a Naturally Rich Country

The image of Indonesia as a naturally rich country is constructed through 17 of 174 rows of concordance from keyword "Indonesia". Table 2 displays the concordance and the collocation that contributes to this meaning.

Table 2. Concordance of Indonesia as a Naturally Rich Country

No	Concordance
1	history and ethnic diversity. **Indonesia** has a well-preserved natural ecosystem with
2	unemployment rate was 5.28%.[162] **Indonesia** has abundant natural resources like oil
3	for Maritime and Investments Affairs, **Indonesia** has 17,504 islands (16,056 of which are
4	the Orangutan wildlife reserve. Moreover, **Indonesia** has one of the world's longets
5	population and densely populated regions, **Indonesia** has vast areas of wilderness that
6	Indonesia and Flora of **Indonesia** Indonesia's size, tropical climate, and archipelagic
7	algal beds, and small island ecosystems.[14] **Indonesia** is one of Coral Triangle
8	for human habitation and agriculture. **Indonesia** is second only to Australia in terms
9	more than 1,650 species in eastern **Indonesia** only.[93] British naturalist Alfred Russel
10	wet and dry seasons, the basis of **Indonesia**'s agricultural system.[106]It would also
11	warm waters that constitute 81% of **Indonesia**'s area ensure that temperatures on land
12	(Wallace Line) between the distribution of **Indonesia**'s Asian and Australasian species
13	(16,024 ft),Puncak Jaya is **Indonesia**'s highest peak, and Lake Toba
14	endemic.[92]Tropical seas surround **Indonesia**'s 80,000 kilometres (50,000 miles) of
15	rain forests that stretch over about 57% of **Indonesia**'s land (225 million acres). Forests
17	an area of 1,145 km2 (442 sq mi).**Indonesia**'s largest rivers are in Kalimantan
18	and Flora of Indonesia **Indonesia**'s size, tropical climate, and archipelagic geography

All data in Table 2 contribute to form the image of Indonesia as a country with rich natural resources. The natural resources owned by Indonesia are *well-preserved natural ecosystem, abundant natural resources, oil, thousands of islands, wildlife reserve, vast areas of wilderness, flora, coral triangle, agriculture, thousands species, seasons, waters, peak, lake, seas, rain forests, rivers, tropical climate*, and *archipelagic geography*.

The collocation which occurs along together the keyword "Indonesia" tends to form the image of Indonesia as a country with abundant natural resources in land, sea, and air. Natural resources are also available in huge quantities. The quantity is represented in number or statistics (such as in data number 3, 9, 13, 14, 15, and 17) or adjective (such as in data number 2, 4, 5, 17). It means that the natural resources owned by Indonesia are something extraordinary. Indonesia is too wealthy or natural resources that it can be described both quantitively and qualitatively. This also leads to convey a positive image to Indonesia for the readers of Wikipedia. This positive image can lead to the free promotion for international citizen to visit Indonesia and enjoy the natural beauty in many forms such as rivers, beaches, natural parks, warm climate, and so on.

Indonesia is Actively Participating in International Sport Competition

Indonesia seems enthusiastic to join in international sport competition. The spirit is represented via this concordance in the table below.

Table 3. Indonesia is Actively Participating in International Sport Competition

No	Concordance
1	basketball, which has a long history in **Indonesia** and was part of the first National Games
2	the Southeast Asian Games, with **Indonesia** appearing as one of the leading competitors.
3	football are the most popular sports. **Indonesia** is among the only five countries that
4	bronze medal in the 1958 Asian Games. **Indonesia**'s first appearance in the AFC
5	sport that contributes the most to **Indonesia**'s Olympic medal tally. Liga 1
6	On the continental level, **Indonesia** won the bronze medal in the 1958 Asian Games
7	On the international stage,**Indonesia** has experienced limited success despite being the first Asian team
8	the leading competitors. In Southeast Asia, **Indonesia** is one of the top sports powerhouses

Indonesia is internationally acknowledged by its achievements in international sport competitions. It develops its sports competition from national to international levels such as *Southeast Asian Games, Asian Games, Asian Football Confederation Cup*, and *Olympic*. The achievements of Indonesia in this field might contribute a positive image to the reputation of Indonesia. People from outside Indonesia might know Indonesia from sports achievements. It also raises the dignity of Indonesia in international forum.

Indonesia is Developing Economy

Indonesia as a developing country is represented in Wikipedia article via the concordance in the table 4. Based on the concordance above, Indonesia is developing its economy via *export and import activities, domestic consumption,* and *international tourism*. It encourages Indonesia to gain income and economic growth as well as to overcome the economy crisis. What makes Indonesia's economy develop is the domestic consumption. Since Indonesia is the fourth country based on the number of its population, domestic or foreign potential investors might find Indonesia is a prospective country to be the target market of their products or services.

Table 4. Concordance of Indonesia is Developing Economy

No	Concordance
1	and Economic history of Indonesia **Indonesia** has a mixed economy in
2	US$19.7 billion to GDP in 2019. In 2018, **Indonesia** received 15.8 million visitors, a growth of 12.5%
3	India, Malaysia, South Korea and Thailand are **Indonesia**'s principal export markets and import partners.
4	of substantial economic growth. **Indonesia** was the country hardest hit by the 1997 Asian financial crisis
5	domestic consumption, helping **Indonesia** weather the 2008-2009 Great Recession.[160]
6	rebellion against it. The economy of **Indonesia** is the world's 16th largest by nominal GDP
7	Indonesia. Since 2011, Wonderful **Indonesia** has been the slogan of the country's international marketing campaign

Indonesia Has a Long History

As discussed in the previous finding, the cultural and natural resources owned by Indonesia are gained by a long process of history. The representation of Indonesia's long history is shown in Table 5.

Table 5. Concordance of Indonesia Has a Long History

No	Concordance
1	Main article: History of **Indonesia** Early history Fossilised remains of Homo erectus
2	until 1949 that the Dutch recognised **Indonesia**'s sovereignty following an armed and
3	Etymology Further information: Names of **Indonesia** The name Indonesia derives from Greek
4	in conflicts among other countries.[126] **Indonesia** was a significant battleground during the Cold War
5	predating the formation of **independent** Indonesia.[15]In 1850,George Windsor Earl
6	officially the Republic of **Indonesia** (Indonesian:Republik Indonesia
7	increasingly powerful Communist Party of **Indonesia**(PKI).[50] Tensions between the military and the PKI

From the concordance above, the history of Indonesia stretches from the era of *Homo erectus* until the internal conflict. It involves the *origin of Indonesia's name, the formation of independence*, and the era of the *Cold War*. This history makes Indonesia as it is now. The

history recorded in this concordance only portrays some important moments such as the independent day and the tension between the communist party and military that caused a horrific tragedy. From this snippet in the concordance line, we could understand why there is no communist party since then. This also indicates communism ideology is not suitable for Indonesia. It might contribute a positive image that Indonesia is a democratic country.

Indonesia is Facing Some Problems

Aside from the two positive images described above, a negative image is also found. It is constructed by the collocation that contribute to conveying meaning that Indonesia faces soma problems. It is presented in Table 6 as follows.

Table 6. Concordance of Indonesia is Facing Some Problems

No	Concordance
1	deaths in 2015).Nevertheless, **Indonesia** continues to face challenges that include maternal
2	primary factor behind much of **Indonesia**'s deforestation.[98] While it can generate
3	The report also indicates that **Indonesia**'s performance is generally below average
4	These issues contribute to **Indonesia**'s poor ranking (number 116 out of 180 countries)
5	This situation makes **Indonesia** the world's largest forest-based emitter greenhouses
6	Indonesia Malaysia confrontation, **Indonesia** withdrew from the UN due to the latter'
7	LGBT issues have recently gained attention in **Indonesia**.[264] While homosexuality is

Based on the collocation occur in the concordance of Table 6, we can sum up some problems faced by Indonesia such as *maternal death, deforestation, bad performance (environmental field), poor ranking (environmental), largest forest-based emitter greenhouses (gas), confrontation,* and *LGBT issues*. By the finding of the problems in table above, we can see an irony in Indonesia if we compare it with the finding in Table 2. As mentioned before, Indonesia is a country full of natural resources, and one of them is the forest. Ironically, from the finding in Table 6, we spot that deforestation is an actual problem in Indonesia. This problem causes a bad position for Indonesia in environmental field, whereas Indonesia has good potentials in the natural environment and ecosystem.

In addition to the ironic problem found, it is also notified from the data that the problems faced by Indonesia are running from the past time until the current problem. The past problem is something political, i.e. confrontation with Malaysia, and the current problem is the moral issue like LGBT.

Indonesia is Developing Technology

How Indonesia develops technology is constructed in some data of concordance as illustrated in the table below.

Table 7. Concordance of Indonesia is Developing Technology

No	Concordance
1	Antariksa Nasional, LAPAN). In the 1970s, **Indonesia** became the first developing country to operate a satellite system
2	Like other developing countries, **Indonesia** began development of the Internet
3	-generation jet fighter KAI KF-X.[185] **Indonesia** has a space programme and space agency,
4	Florida, United States.[187]As of 2019, **Indonesia** has launched 18 satellites for various purposes,
5	Aerospace (PT. Dirgantara Indonesia), **Indonesia** has provided components for Boeing
6	and technology in Indonesia **Indonesia**'s expenditure on science and technology is relatively low

The concordance indicates that Indonesia develops technology in the form of *satellite, internet, space program*, and *plane*, although the budget might not be proportional. It supports the previous finding from the economy perspective, Indonesia is a developing country. Although the technology is to meet domestic needs, it is a pride for a developing country like Indonesia. It shows that there is a good will from government and the people to build a better civilization with technology.

Indonesia is Actively Participating in International Affair

The activity of Indonesia in many various international affairs is recorded by Wikipedia and proven via this concordance. Indonesia as a part of international community positively contribute to making the world better via its activism. The foreign policy applied by Indonesia is free and active. Free means it is non-aligned and active means it supports world peace. Indonesia is listed as an active member of *UN, OIC, ASEAN*, and contributors in many *humanitarian and development aids*. This leads to the raising of dignity and the good reputation of Indonesia in international community. Indonesia is well-known as a country with high solidarity to assist the neighborhood countries or those who suffer from natural disaster or war.

Table 8. Concordance of Actively Participating in International Affair

No	Concordance
1	in this achievement.[184] **Indonesia** has also joined the South Korean programme to manufacture the fifth-generation jet fighter KAI KF-X
2	member state had attempted a withdrawal.[134] **Indonesia** has been a humanitarian and development aid
3	has ties with Israel, albeit discreetly.[131] **Indonesia** has been a member of the United Nations
4	Foreign relations of **Indonesia** Indonesia maintains 132 diplomatic missions abroad
5	Organisation of Islamic Cooperation(OIC).[132]**Indonesia** is a signatory to the ASEAN
6	maintaining a non-aligned stance has characterised **Indonesia**'s foreign policy since then

IV. Conclusion

Based on the corpus-based discourse analysis applied in the article entitled Indonesia in Wikipedia, it is objectively constructed in the readers' mind that Indonesia is represented as a country with rich cultural and natural resources; actively join in both international sports competition and international affair. Despite common problems faced by many countries, Indonesia is developing its economy and technology as well. Commonly, the image of Indonesia in this article tends to be positive.

This study implies that Wikipedia is objective in describing Indonesia. Not only does Wikipedia write the positive side of Indonesia, but also it elaborates the problems in Indonesia. By applying corpus-based discourse analysis, we can find that Wikipedia records the issues developing in Indonesia, from political issue to the most contemporary moral issue like LGBT.

Based on the implication, it is suggestable that other researchers could conduct research related to the other specific and contemporary issues in Indonesia using corpus-based analysis to discover the discourse recorded in society via electronic texts such as sites, social media, and electronic document releases.

References

[1] Al Fajri, M. S. (2017). Hegemonic and minority discourses around immigrants: A corpus-based critical discourse analysis. *Indonesian Journal of Applied Linguistics*, *7*(2), 381–390. https://doi.org/10.17509/ijal.v7i2.8349

[2] Baker, P. (2008). Using Corpora in Discourse Analysis. London: Continuum.

[3] Baker, P. (2010). Corpus Method in Linguistics in L. Litosseliti (ed), *Research Methods in Linguistics* (pp.93-113). London: Continuum.

[4]   Gabrielatos, C. (2014). *Department of Discourse & Communication Studies Corpus Approaches to Discourse Studies The basics Costas Gabrielatos Edge Hill University*. *September*. https://doi.org/10.13140/2.1.2410.3683

[5]   Gabrilovich, E., & Markovitch, S. (2007). Computing semantic relatedness using wikipedia-based explicit semantic analysis. *IJCAI International Joint Conference on Artificial Intelligence*, *January 2015*, 1606–1611.

[6]   Hunston, S. (2002). Corpora in Applied Linguistics. Camdridge: Cambridge University Press.

[7]   Kittur, A., Chi, E. H., Pendleton, B. A., Suh, B., & Mytkowicz, T. (2007). Power of the few vs. wisdom of the crowd: Wikipedia and the rise of the bourgeoisie. *25th Annual ACM Conference on Human Factors in Computing Systems*, 1–9.

[8]   Sunarsih, N. (2017). Pola Representasi Pihak Israel Dalam Ensiklopedia Britannica. *Kandai*, *13*(2), 173. https://doi.org/10.26499/jk.v13i2.191

[9]   Zesch, T., Müller, C., & Gurevych, I. (2008). Extracting lexical semantic knowledge from Wikipedia and Wiktionary. *Proceedings of the 6th International Conference on Language Resources and Evaluation, LREC 2008*, 1646–1652.