Designing a Chatbot Application for Student Information Centers on Telegram Messenger Using Fulltext Search Boolean Mode

Hokianto Candra[1)*], Rino[2)], Riki[3)]

[1)2)3)]*Buddhi Dharma University*
Jalan Imam Bonjol No 41, Tangerang, Indonesia

[2)] rino@ubd.ac.id

Abstract

Chatbot can be used as a solution of human resource issues in term of answering user's questions about STMIK PGRI Tangerang. A chatbot application can have a conversation or answer user's questions automatically. Knowledge is needed for chatbot to work properly. The general scheme of chatbot's knowledge is the existence of patterns and responses. This research will implement the Fulltext Search Boolean Mode for the pattern-matching process with the input received by the chatbot application based on Telegram Messenger Bot. Chatbot application which created in this research can provide answers to information that is relevant to the input or questions provided by users.

I. INTRODUCTION

Information can be obtained from various existing internet services using various devices. A research conducted by APJII (Asosiasi Penyelenggara Jasa Internet Indonesia) in 2017 [1] shows that smartphones are the most widely used device to access the internet, and the most accessed service is chat (instant messaging) service [1].

STMIK PGRI Tangerang has information media for its students and the general public, namely the website. But sometimes, the information provided on the website is incomplete, so users will ask their questions directly via telephone or social media. The asked questions are often similar or have been already answered on the website's FAQ (Frequently Asked Question). The questions received can't be answered immediately because of limited human resources such as speed and servicing times.

Chat applications have now been pinned Bot as the technological development, one of which is Telegram Messenger. Bot is a program that runs in certain environments to do automated jobs, one of type is ChatBot. ChatBot is a computer program that can carry out conversations through writing or chat media [2].

The general scheme of ChatBot is patterns and templates (responses). The pattern-matching can be done by various methods, one of them using Fulltext Search Boolean Mode from MySQL. In addition to being able to provide the most relevant template for input, applying this method can also reduce the number of records in the database.

II. METHODS

Telegram Bot API

According to [3] Telegram Bot API (Application Programming Interface) is a open source technology provided by Telegram to build Telegram Bot applications for developers. This Bot API is an HTTP-based interface for connecting Bots developed by developers with Telegram system.

Telegram Bot is a special account that functions as an interface to run the program code that has been created. This account can be created through a special Telegram Bot named @BotFather [4].

[*] Corresponding author

According to [5] Telegram Messenger offers capability of synchronization, super fast service, reliable backup and better security features.

There are several methods used to design a Bot in Telegram Messenger. Some of them are sendMessage, sendPhoto, sendDocument, sendVideo, sendLocation, editMessageText [6]. All queries to the Telegram Bot API must go through HTTPS and be presented in the form https://api.telegram.org/bot<token>/METHODS_NAME. The Telegram Bot also supports sending formatted text like bolded text, italic text or URLs, using the syntax below.

```
*bold text*
_italic text_
[inline URL](http://www.example.com/)
[inline mention of a user](tg://user?id=123456789)
`inline fixed-width code`
```block_language
pre-formatted fixed-width code block
```
```

There are two ways to run a Telegram Bot, Long-Polling and Webhook [7]. Long-Polling is the default method in making a Telegram Bot, where the client makes a continuous requests for updates to the server in periodical time interval. The advantages of this method are that it can use any computer as a client that will make requests to the API or Telegram Bot server. While the weakness of the Long-Polling method is the request-response process and can consume large resources, and the possibility of asynchronous data between client and server resulting in long response or unreal-time data exchange.

The advantages of the Webhook method are the real-time data exchange and only need to configure the callback URL, then the sending server will send data to the configured URL when there is an update. And the weakness of the webhook method is the need for a server that has an HTTPS connection, or uses hosting if it doesn't exist. The result of research conducted by [7] show that the Long-Polling method has a higher response time (slower) than the Webhook method.

Fulltext Search Boolean Mode

According to [2] said that "Fulltext Search Boolean Mode is a feature used for the matching process. Matching is different from ordinary matching. Matching with this method can produce the most relevant results. This feature can be applied to the InnODB and MyISAM engines in MySQL version 5.6. The advantage of Fulltext Search Boolean Mode is that it can implement operators in search".

The use of Fulltext Search Boolean Mode is accomplished by first adding FULLTEXT index on the column that will be used as a search criterion. The query below is an example of making a table that contains the FULLTEXT index in the pattern column.

```
CREATE TABLE `knowledge` (
 `id` int AUTO_INCREMENT,
 `pattern` varchar(255),
 PRIMARY KEY (`id`),
 FULLTEXT KEY `pattern` (`pattern`)
)      ENGINE=InnoDB      DEFAULT
CHARSET=latin1
```

Then the column that contains the FULLTEXT index can be used in Fulltext Search Boolean Mode search / matching. The query below is an example of using Fulltext Search Boolean Mode in searching for the word "lokasi unud".

```
 SELECT `id`, `pattern`, MATCH(`pattern`)
AGAINST('lokasi STMIK PGRI' IN BOOLEAN
MODE) AS relevansi FROM `knowledge`
ORDER BY(relevansi) DESC;
```

The query will display `id`, `pattern` and `relevance` matching results with input sentences, sorted from the highest relevance value. The Fulltext Search Boolean Mode matching is based on BM25 algorithm and TF-IDF ranking [8]. Boolean Mode function could be seen in equation (1).

$$\${rank\} = \${TF\} * \${IDF\} * \${IDF\} \quad (1)$$
$$\text{or}$$
$$\${rank\} = \${TF\} * (\${IDF\})^2$$

Where:

TF : Term Frequency, number of words appearing in a sentence
IDF : Inverse Document Frequency

The IDF value could be obtained from the number of documents divided by the number of related documents, through equation (2).

$$\${IDF\} = \log_{10}(\frac{\${total_{records}\}}{\${matching_{records}\}}) \quad (2)$$

Equation (1) and equation (2) are used to rank one word. If there are more than word, the rank value is obtained by adding up the rank values of each word. It can be written in equation (3).

$$\${rank\} = \sum(\${TF\} * \${IDF\} * \${IDF\}) \quad (3)$$

Regular Expression

According to [9] Regular Expression is a specific kind of text pattern that can be used with many modern applications and programming languages. Regular Expression can be used to verify whether input fits into the text pattern, to find text that matches the pattern within a larger body of text, to replace text matching the pattern with other text or rearranged bits of the matched text, to split a block of text into a list of subtexts.

methods sections often come disguised with other article-specific section titles, but serve a unified purpose: to detail the methods used in an objective manner without introduction of interpretation or opinion. The methods sections should tell the reader clearly how the results were obtained. They should be specific. They should also make adequate reference to accepted methods and identify differences.

III. RESULTS

In this research, the pattern matching process using Fulltext Search Boolean Mode is done by matching user input with data in the knowledge_base table which has 106 records. Example of input that will be matched is "dimana alamat STMIK PGRI?" by using a query "**SELECT `id_kb`, `pattern`, `template`, `flag`, MATCH(`pattern`) AGAINST("dimana alamat STMIK PGRI?" IN BOOLEAN MODE) as `relevansi` FROM `knowledge_base` WHERE MATCH(`pattern`) AGAINST("dimana alamat STMIK PGRI?" IN BOOLEAN MODE) ORDER BY(relevansi) DESC**". The results of running the query can be seen in the figure below.

Fig. 1 Results for Running a Fulltext Search Boolean Mode Matching Query

Figure 1 shows that the record with `id_kb`=4 has the highest relevance value of 4.4154582023620605 followed by `id_kb`=65 with relevance value of 2.9731273651123047 and `id_kb`=5 with relevance of 1.4423308372497559. The manual calculations will be performed on all the three records.

Table 1. Calculation of Term Frequency (TF)

id_kb	pattern	Word	TF
4	alamat lokasi letak kampus stmik pgri tangerang	dimana	0
		alamat	1
		STMIK	1
		PGRI	1
65	alamat	dimana	0
		alamat	1
		STMIK	0
		PGRI	0
5	kontak nomor telepon telpon fax stmik pgri tangerang	dimana	0
		alamat	0
		STMIK	1
		PGRI	1

Table 2. Perhitungan Inverse Document Frequency (IDF)

Word	Matching Record	Total Record	IDF
dimana	0		$\log_{10}\left(\frac{106}{0}\right)$
alamat	2	106	$\log_{10}\left(\frac{106}{2}\right)$
STMIK	15		$\log_{10}\left(\frac{106}{15}\right)$
PGRI	15		$\log_{10}\left(\frac{106}{15}\right)$

Table 3. Calculation Results for the Relevance Value of Each Record

id_kb	Word	Formula	Temporary Result	Final Result
4	dimana	$0 * \left(\log_{10}\left(\frac{106}{0}\right)\right)^2$	0	4.4154581691461000
	alamat	$1 * \left(\log_{10}\left(\frac{106}{2}\right)\right)^2$	2.973127274487557	
	STMIK	$1 * \left(\log_{10}\left(\frac{106}{15}\right)\right)^2$	0.7211654473292715	
	PGRI	$1 * \left(\log_{10}\left(\frac{106}{15}\right)\right)^2$	0.7211654473292715	
65	dimana	$0 * \left(\log_{10}\left(\frac{106}{0}\right)\right)^2$	0	2.973127274487557
	alamat	$1 * \left(\log_{10}\left(\frac{106}{2}\right)\right)^2$	2.973127274487557	
	STMIK	$0 * \left(\log_{10}\left(\frac{106}{15}\right)\right)^2$	0	
	PGRI	$0 * \left(\log_{10}\left(\frac{106}{15}\right)\right)^2$	0	
5	dimana	$0 * \left(\log_{10}\left(\frac{106}{0}\right)\right)^2$	0	1.4423308946585430
	alamat	$0 * \left(\log_{10}\left(\frac{106}{2}\right)\right)^2$	0	
	STMIK	$1 * \left(\log_{10}\left(\frac{106}{15}\right)\right)^2$	0.7211654473292715	
	PGRI	$1 * \left(\log_{10}\left(\frac{106}{15}\right)\right)^2$	0.7211654473292715	

Table 4. Results Comparison of Manual Calculations Value and Query Running Value

id_kb	Relevance Result of Manual Calculation	Relevance Result of Running Query	Selisih
4	4.4154581691461000	4.4154582023620605	-0.0000000332159605
65	2.973127274487557	2.9731273651123047	-0.0000000906247477
5	1.4423308946585430	1.4423308372497559	0.0000000574087871

Based on the result of Fulltext Search Boolean Mode matching, the highest relevance value is found in the record with `id_kb`=4 and is the answer that the chatbot application will give to the input "dimana alamat STMIK PGRI?".
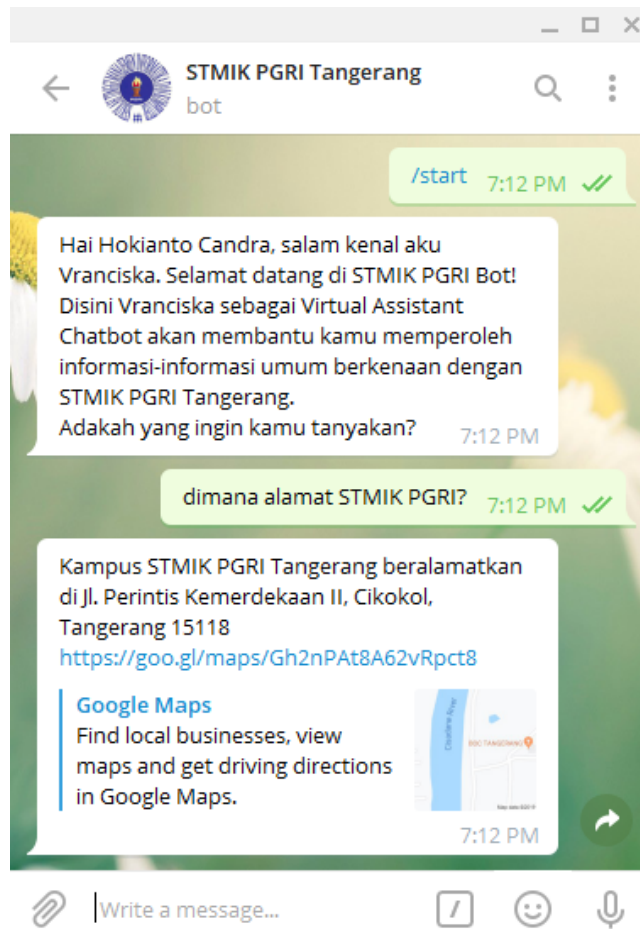

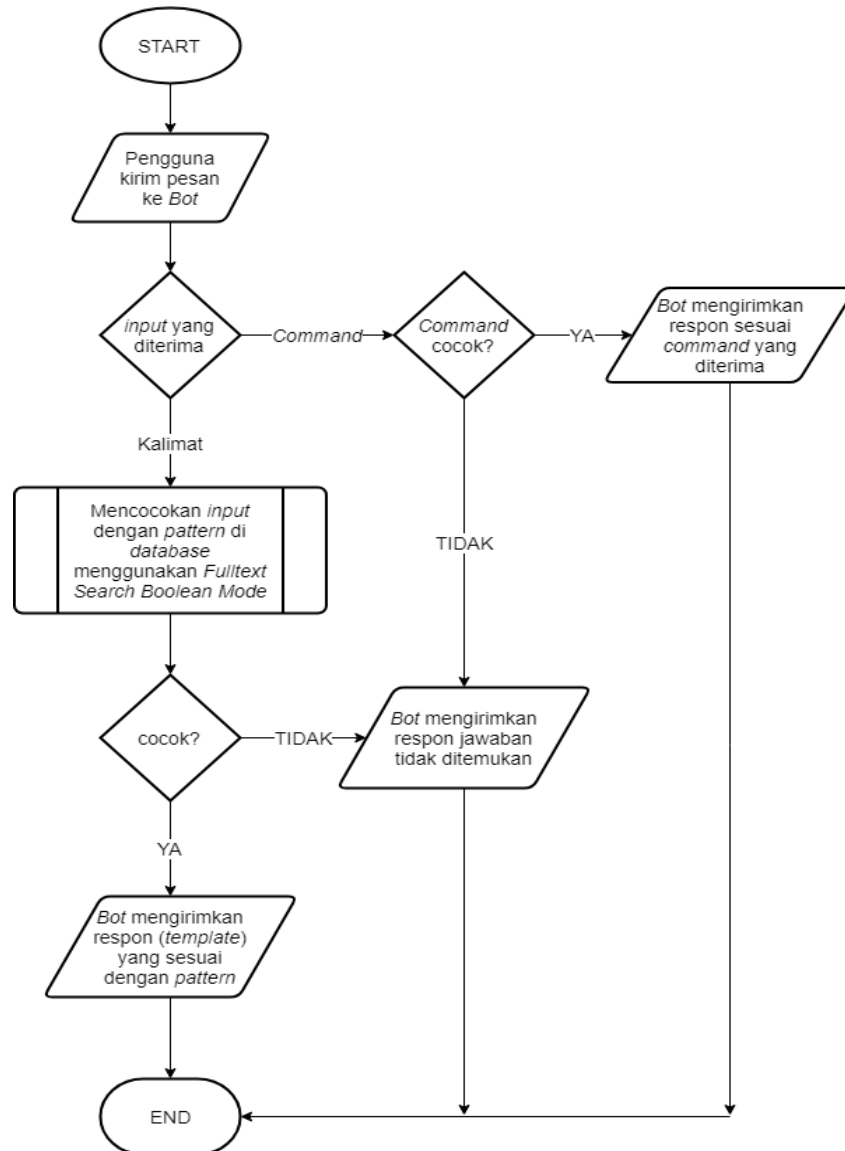Fig. 2 Chatbot's Response to "dimana alamat STMIK PGRI?" Input

IV. DISCUSSION

Fig. 3 Chatbot Application Flowchart

There are two types of input or message that can be received by Bot. To be able to distinguish between the two types of input received, regular expressions are used.

1. Command message (*command*) or can also be called a menu. This message always starts with a slash "/" according to Telegram standards. For example "/start", "/help". For special purposes, developers can create commands with other symbol prefixes such as the exclamation mark "!" or the dollar symbol "$". The regular expression syntax used to find this type of message is as follows.

$$/\verb|^([\!\/\$])([[:word:]]+)(\s?)(.*)|/$$

2. Text message in form of normal conversation text using Indonesian language that is good and right in accordance with KBBI (Indonesia Dictionary). Then the pattern matching process using Fulltext Search Boolean Mode is performed on this text message. The regular expression syntax used to find this type of message is as follow.
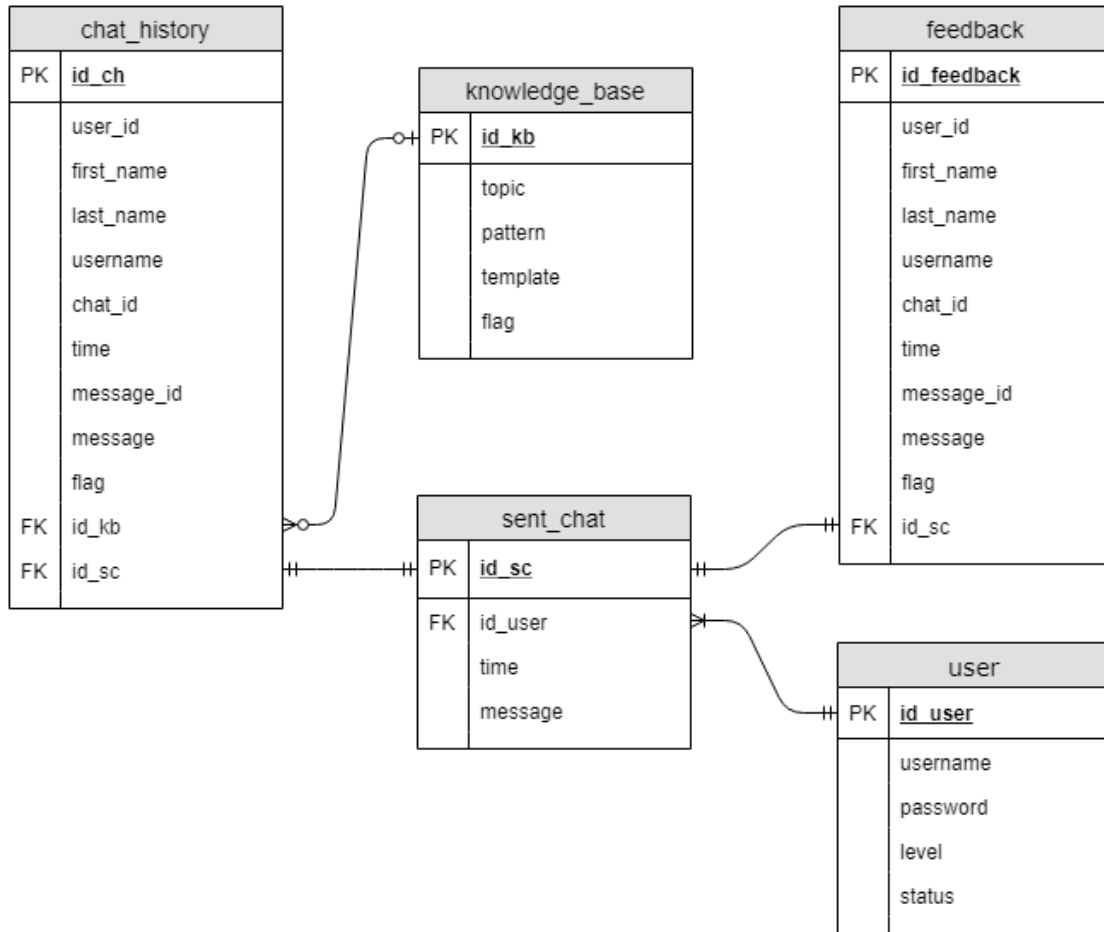
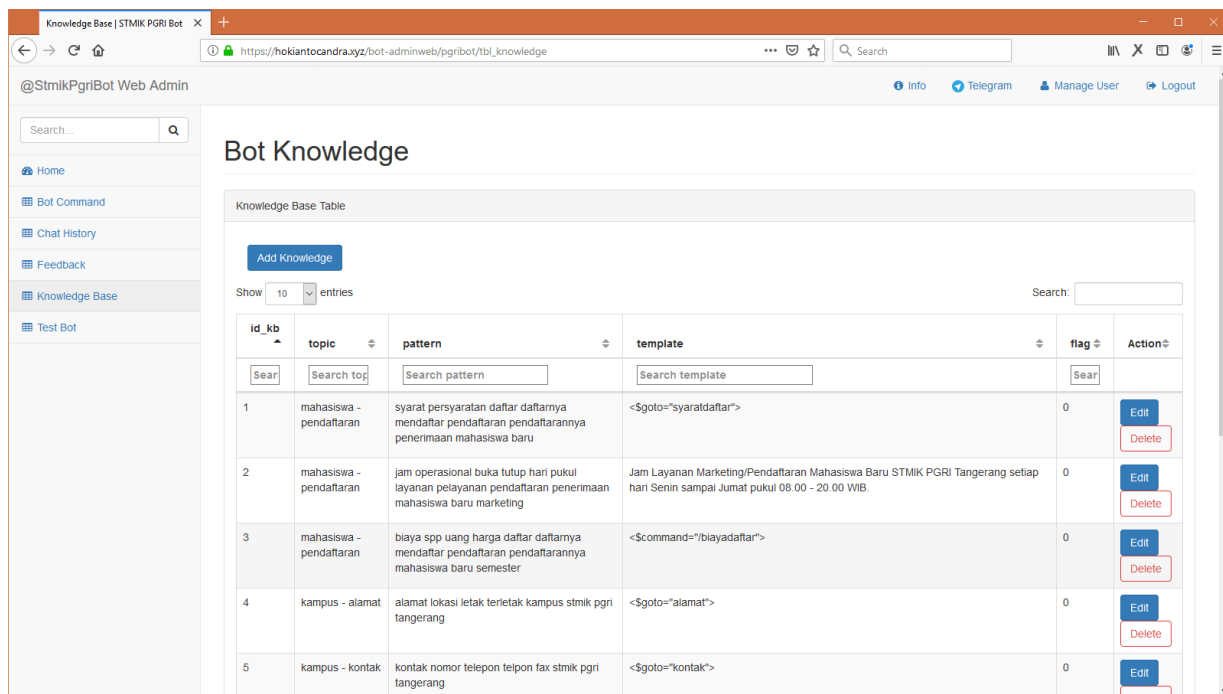$$/\verb|^(\b\w +\b)(\W+)? $|/$$

Fig. 4 Database Relational Schema
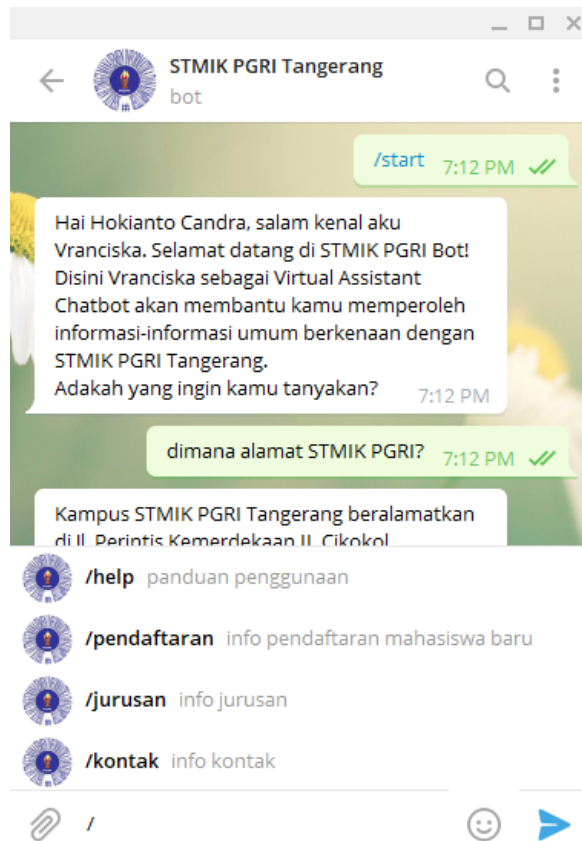

Fig. 5 Bot Knowledge Table Display in Admin Web

Fig. 6 Command/Menu Display in Chatbot

he discussion section is where the article interprets the results to reach its major conclusions. This is also where the author's opinion enters the picture. the discussion is where the argument is made. Common features of the discussion section include comparison between measured and modelled data or comparison among various modelling methods, the results obtained to solve a specific engineering or scientific problem, and further explanation of new and significant findings.

V. CONCLUSIONS

1. With this chatbot application, users (especially students and prospective students) can easily and quickly obtain information about STMIK PGRI Tangerang. This chatbot application can answer user's questions automatically and is available 24 hours.
2. Based on the result of the questionnaire, the responses or answers provided by the chatbot application are precise, consistent, fast, clear and complete in accordance with the statements or questions submitted by users.
3. The implementation of Fulltext Search Boolean Mode in the chatbot application can provide the most relevant response/answer to user's input. The response is obtained from a record with the highest and unique relevance value. The relevance value is influenced by Term Frequency and Inverse Document Frequency. The use of this method can also reduce the number of records in the database because several related search words could be entered in one record. However, in certain cases this method is not appropriate, for example when there are several records with different templates but have some same words in the pattern.

REFERENCES

[1]   R. Riki, A. Hermawan, and Y. Kurnia, "Voice Over Internet Protocol Based Communication Design (VoIP) With 3CXSystemPhone On Android Smartphone," *bit-Tech*, vol. 1, no. 1, pp. 1–8, Aug. 2018.

[1]   APJII, "Asosiasi Penyelenggara Jasa Internet Indonesia," 2017. [Online]. Available: https://apjii.or.id/survei2017. [Accessed 28 July 2019].

[2]   I. N. S. Paliwahet, I. M. Sukarsa and I. K. G. D. Putra, "Pencarian Informasi Wisata Daerah Bali menggunakan Teknologi Chatbot," *LONTAR KOMPUTER,* vol. 8, no. 3, pp. 144-153, Desember 2017.

[3]   R. D. Risanty and A. Sopiyan, "Pembuatan Aplikasi Kuesioner Evaluasi Belajar Mengajar Menggunakan Bot Telegram pada Fakultas Teknik Universitas Muhammadiyah Jakarta (FT-UMJ) dengan Metode Polling," pp. 1-9, 2017.

[4]   Telegram, "Bots: An Introduction for developers," Telegram, [Online]. Available: https://core.telegram.org/bots. [Accessed 24 July 2019].

[5]   T. Sutikno, L. Handayani, D. Stiawan, M. Agus Riyadi and I. Much Ibnu Subroto, "WhatsApp, Viber and Telegram: which is the Best for Instant Messaging?," *International Journal of Electrical and Computer Engineering (IJECE),* vol. 6, no. 3, pp. 909-914, Juni 2016.

[6]   Telegram, "Telegram Bot API," [Online]. Available: https://core.telegram.org/bots/api. [Accessed 24 July 2019].

[7]   H. Soeroso, A. Z. Arfianto, N. E. Mayangsari and M. Taali, "Penggunaan Bot Telegram Sebagai Announcement System pada Intansi Pendidikan," *Seminar MASTER 2017 PPNS,* vol. 2, no. 1, pp. 45-48, Desember 2017.

[8]   "MySQL :: MySQL 8.0 Reference Manual :: 12.9.2 Boolean Full-Text Searches," Oracle, [Online]. Available: https://dev.mysql.com/doc/refman/8.0/en/fulltext-boolean.html. [Accessed 24 July 2019].

[9]   S. Levithan and J. Goyvaerts, Regular Expressions Cookbook: Detailed Solutions in Eight Programming Languages, 2nd ed., Sebastopol, California: O'Reilly Media, 2012.