International Journal on Advanced Science Engineering Information Technology

Vol.10 (2020) No. 5
ISSN: 2088-5334

Human-Robot Interaction Based on Dialog Management Using Sentence Similarity Comparison Method

Dinda Ayu Permatasari[a,1], Hanif Fakhrurroja[a,2], Carmadi Machbub[a,3]

[a]*School of Electrical Engineering and Informatics, Bandung Institute of Technology, Jl. Ganesha 10, Bandung, 40132, Indonesia*
E-mail: [1]prmtsrdinda@gmail.com; [2]hani002@lipi.go.id; [3]carmadi@lskk.ee.itb.ac.id

***Abstract*—Advances in developing dialogue systems regarding speech recognition, language understanding, and speech synthesis. Dialogue systems to support human interaction with a robot efficiently by using spoken language. Facilities that provide convenience in carrying out daily activities for someone, such as older people, are necessary. The existence of Human-Robot Interaction (HRI), so that this interaction can give orders to the robot to do work that cannot be done by humans. This study presents a dialogue management system for HRI with a comparison sentence similarity method between TF-IDF (Term Frequency-Inverse Document Frequency) Cosine Similarity Algorithm and Jaccard Coefficient and using Finite State Machine (FSM). Dialogue Management is a way to find the response of the answer. When the user says something or in other words, is responsible for managing the flow of the conversation to command the robot. TF-IDF is used to give the weight of the term relationship and comparison between Cosine Similarity and Jaccard Coefficient for comparison method to determine the classification of similarity sentences from the dialogue manager to improve the intent of the dialogue, for the FSM method to set the sequence flow dialogue. We use Google Cloud Speech API as an engine for speech to text using Kinect V2 as an audio sensor. There are eight scenarios created in this system. The speech recognition process using Google Speech for an average of 2.62 seconds shows a reasonably fast response. TF-IDF Cosine Similarity method can produce enough accuracy of 97.43%, and Jaccard Coefficient indicates an accuracy level of 91.57%. The state of the FSM method can be considered as an efficient structure for building dialogue management.**

***Keywords*— dialogue manager; TF-IDF; cosine similarity; finite state machine; human-robot interaction; Google cloud speech.**

I. INTRODUCTION

Dialogue is a way of communication that became a basis of interaction between humans. Dialogue also makes it easy for us to understand the intent of talking to someone who is communicating with the other person. The dialogue system is communication between humans as users and machines with natural language [1]. Dialogue is a process part of the Human-Robot Interaction (HRI), immensely to help the elderly and people who have physical limitations. Dialogue management is essential for a dialogue system to determine the flow of communication between users and robots. The dialogue management system allows users to interact with natural language through spoken speech. Challenges in natural language processing (NLP), namely the relationship between speech and the intent of the user towards user desires.

To build a dialogue management system is a challenging work in automatic speech recognition (ASR) and technology for an understanding speech from users. The system for understanding words in the flow of the dialogue refers to artificial intelligence (AI) with a machine learning approach that depends on information or data [2]. Classification for dialogue where the classification decision to understand the intent is very important in understanding the user's speech in dialogue [3].

The dialogue system with natural language has been developed in recent years. Dialogue is used as a home assistant for people with disabilities and the elderly by building dialogue modelling methodologies based on various models and descriptions of the entire system [4]. The development of a system dialogue provides information to the elderly using classification that does not involve language modeling, which usually uses sub-lexical features [5]. The method used is a classification of actions based on word order and analysis of meaning using CVM sequences. The result shows that sub-lexical classification is better than HMM [5]. Hybrid management dialogue is also done as a solution model for complex dialogue sequences, with subsequent dialogues using the Java Rule Engine (JESS) and Partially Observable Markov Decision Process (POMDP) for solutions to complex dialogue decision-making models [6]. Dialogue with more tasks and more in-depth dialogue is based on the standard situation for the home environment for interaction of humanoid robot control techniques [7].

Another research is combining a dialogue manager with intelligent robots using the FSM method to determine the state of the next robot and produce intelligent robot services [8]. Others using Finite-State Turn-Taking Machine (FSTTM) to control the behavior of conversation, and the result shows that this method gives a high response than previous approaches [9]. Natural Language Understanding using TF-IDF vector and linear SVM classification results in information retrieval methods showing better performance than NN based models. The system used to handle user requests with FSM with predetermined modeling [10]. TF-IDF weights, based on the similarity of the sentence answers that match the highest, obtain an index of information showing the results of the dialogue system for information retrieval with an average accuracy rate of 84.33% [11].

In the research that has been carried out to develop natural language processing, one of them is management dialogue that is carried out like chatbot, which uses machine learning and artificial intelligence, with the appearance of a platform that can attract people to be able to interact. For example, the EU project developed a new technology for the dialogue system called ALICE by combining speech with graphics into the car [12].

Previously we researched in management dialogue with Artificial Intelligence Markup Language (AIML) method. AIML can characterize data object types and draw branching from the programs it processes by forming user input patterns and dialogue responses based on the basic unit of dialogue. The most important objects in AIML are category tags that explain the knowledge units of the conversation. These pattern tags identify input from users and template tags that respond to specific input from users. That method shows an accuracy rate of 92.4%, which is quite sufficient for manager dialogue but not too flexible to understand the intent of the user because it must always follow the tag pattern that has been built [13].

However, from these results, we still want a dialogue management system that can produce a high level of accuracy and flexible enough to understand the intent of the dialogue from the user. The development is carried out to be able to produce a management dialogue with the intent that can be well understood with high accuracy for management dialogue for the case for robot assistants in the home environment for the elderly. With the challenge of classifying the intent of the user, the underlying framework is needed.

In this study, we implemented the TF-IDF method to give the weight of the given word to the database. We used this algorithm because it is the most efficient computational algorithm for searching documents that are the same as queries [10] and produce useful accuracy sentence classification [14]. On the other hand, the cosine similarity method is used for calculating the similarity of two vectors using keywords in the database. Jaccard coefficient compares similarities and diversity in a sample set of dialogue sentences [11]. Cosine similarity and Jaccard coefficient were compared to see the best results accuracy for classifying based on sentences like datasets.

The similarities in data mining are usually described in the distance with dimensions that include the features of the sentence in the dialogue. If the distance is small, then there is a high level of similarity to be classified based on class. However, if the distance is considerable, the similarity level is low. And also, to control the flow of dialogue using the finite state machine, so the system can understand the use of fixed rules for taking appropriate actions for this conversation. So that from the dialogue system that is created, this can control the Bioloid Robot to do the commands given.

II. Materials and Method

A. Overview of the Proposed System

Figure 1 illustrates the overall architecture system proposed for management dialogue, and a personal computer is used to process speech recognition and dialogue management system. Google Cloud Speech engine is used for the speech recognition process that is integrated into Microsoft Visual Studio 2017 software. The sound sensor used, as an audio sensor, is Microsoft Kinect 2.0 microphone array component. Term Frequency-Inverse Document Frequency (TF-IDF) is a method used to give weight to the relationship of a term to a sentence. After the vectorization and weighting process with TF-IDF, sentence similarity scores were calculated using the Cosine Similarity algorithm. The answer to the question taking from the highest angle of cosine. But the flow of the dialogue process that is answer and question is controlled using the finite state machine with eight dialogue scenarios. There are two output responses from this system, namely the audio response generator to provide feedback to the user through the sound of the robot speaker and also provide an output in the form of a command to the Bioloid Premium Robot in the case in this paper to show the object desired by the user.

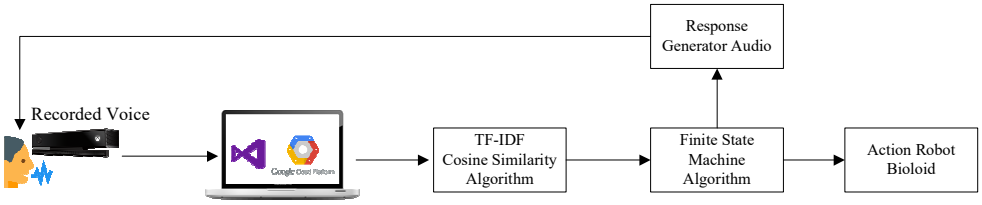

Fig. 1 Architecture system of Dialogue Management

B. Google Cloud Speech API

Speech recognition is the process of converting human voice signals into written language. Speech recognition is the most important part used in this dialogue management system because reliable speech recognition can make dialogue by the scenario. Cloud API development has been growing, one of which is the Google Cloud Speech API,

which currently has 120 languages. Google Cloud Speech API (Application Programming Interface) is one of the speeches to text services provided by the largest search engine, Google, which can be integrated by the developer into the application. Cloud API can determine which application software to use can interact with cloud-based platforms [15].

Google Cloud Speech API service helps recognize sounds from audio data sent through requests and unites them to voice storage on Google Cloud storage. Automatic speech recognition of Google Cloud Speech implements deep learning neural network algorithms for the audio sent by users with high accuracy [16].

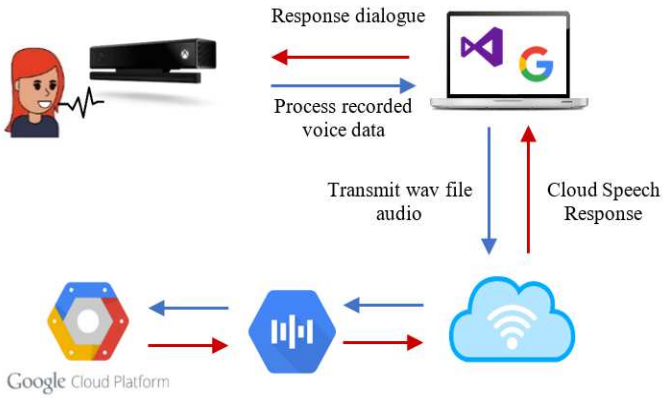

Fig. 2 Architecture Speech Recognition Using Google Cloud Speech

The method used by Google Cloud Speech API is synchronous recognition by sending audio data in the form of a wav file to the Speech-to-Text API, and introducing the data and returning the results after all audio is processed [16]. Figure 2 shows the speech recognition architecture using the Google Cloud Speech API, then processed recorded voice data through the software on a personal computer (PC). Sound files are recorded in real-time, then transmitted to Google Cloud Server, after the Google Cloud Speech Platform recognizes sound after receiving a sound package then sends the converted text back to the user [17].

In the previous research, Google used development for Listen, Attend, and Spell (LAS) model for acoustics, pronunciation, and language for ASR (Automatic Speech Recognition). The component of ASR (Automatic Speech Recognition) is a basic structure that indicates words from the model pieces that are used as substitutes for grapheme. [I1] The structure improvements model sequence to sequence has used graphene as a separate language output unit as AM, PM, and LM into one neural network [18].

C. Classification Sentence Similarity

In this case, the TF-IDF algorithm is a method to give the weight of the term relationship with the sentences. This algorithm is a weighting scheme that can be categorized as a statistical procedure, even though the results are deterministic. In Figure 3, there are preprocessing data, which is the first process applied to the text to convert into numerical data. Tokenizing is the process of breaking sentences into groups of words and separating each space, after which stopwords, by removing non-essential words by checking the words that are parsed. Sometimes a few common words will appear even if they are of little value, so the stopwords function to select the appropriate documents according to user needs [19]. So, the input sentence can be vectorized based on a bag of words. In the training process, weighting adjustments for sentences in the database are given to each element in the bag of a word and weighted by inverse document frequency. After the sentence element has a weight, the sentence similarity score is calculated with each vector, and the validation answer is obtained from the highest angle value of the cosine [11].

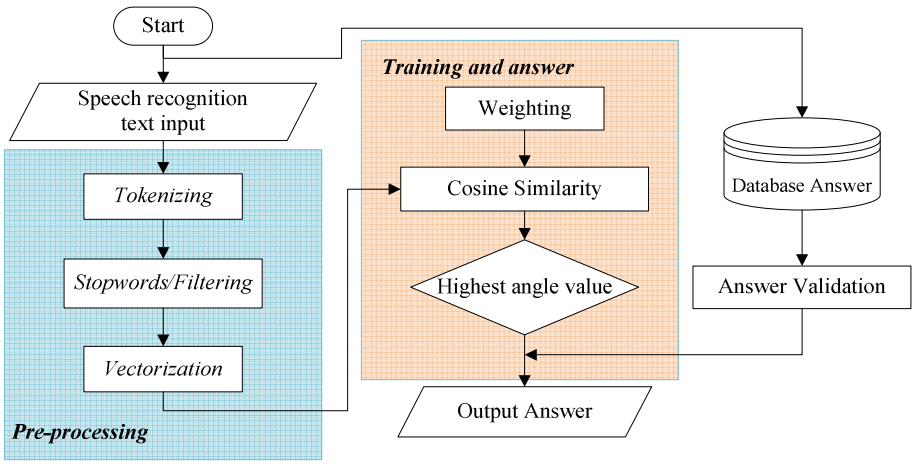

Fig. 3 TF-IDF Cosine Similarity Algorithm

1) TF-IDF Algorithm

TF-IDF serves to classify input sentences into one class and retrieve the answer information from the dataset by vectorizing all sentence entries. So, tf calculates the occurrence of the term in the sentence. The weight of the term in the sentence vector is the local parameter and the global parameter. tf is a local parameter that is a frequency of the term t in the sentence d with $f(d_j, t_i)$, meaning that the frequency of occurrence of the t term to i in the sentence d to j [11].

$$tf(t_i, d_j) = f(d_j, t_i) \quad (1)$$

The global parameter is idf which is the inverse frequency of sentences that idf calculates the appearance of the term in the dialog database, where D is a set of sentences of scenario dialog.

$$idf(t_i, \mathrm{D}) = 1 + log\,\frac{D}{d(t_i)} \quad (2)$$

Hence, the equation of TF-IDF is used to weight the document to d then to the term t.

$$w(t_i, d_j) = tf(t_i, d_j) * idf(t_i, D) \quad (3)$$

This term frequency is usually to increase recall in information retrieval but cannot be ascertained to improve precision.

2) Cosine Similarity

The way to classify the same two sentences by considering the magnitude of the vector difference between two sentences. It is assumed that these two vectors are written A and B to be like eq. (4)

$$A = (A_1, A_2, \dots, A_n)$$
$$B = (B_1, B, \dots, B_n) \quad (4)$$

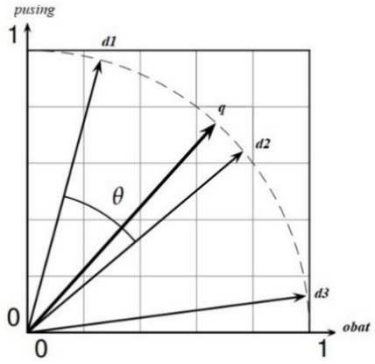

Fig. 4 Illustration of Cosine Similarity

A set of sentences in a collection can be seen as a set of vectors in a vector space, where there is one axis for each term, as can be seen in figure 4. Compensate the effects of sentence length, a way to quantifying the similarity between two sentences by computing the cosine similarity of the representations of vectors A and B [19].

$$\cos\theta = \frac{A \cdot B}{|A||B|} = \frac{\sum_{i=1}^{n} A_i \,.\, B_i}{\sqrt{\sum_{i=1}^{n}(A_i)^2} \cdot \sqrt{\sum_{i=1}^{n}(B_i)^2}} \quad (5)$$

Where, the inner product of vectors A and B is calculated, which is considered a unit normalized length, while the denominator is a product of their euclidean length. In this cosine idea, what is important is the angle θ between two vectors, shown in figure 4. Indeed, this cosine calculation cannot produce negative elements in the raw vector frequency because, in the vector, there are only positive integers. So the equation can be explained again as follows.

$$\cos(q, d) = \frac{\sum(w_{q,t} \cdot w_{d,t})}{W_d} \quad (6)$$

The translation of the equation becomes equation (7), (8), and (9). Eq. (7) is based on equation (1) calculation that the frequency of term t obtained from sentences $\boldsymbol{S}$ and Eq. (8) based on eq. (3) calculation of the weight from inverse document frequency to retrieve sentences.

$$W_d = \sqrt{\sum_{t \in d} w_{d,t}^2} \quad (7)$$

$$w_{d,t} = \log(f_{d,t} + 1) \quad (8)$$

$$w_{q,t} = \log(f_{q,t} + 1) \cdot \log(D/f + 1) \quad (9)$$

A sentence is classified according to the highest rank by cosine if it contains many query terms and if the words are common in sentences but are relatively rare in collections. Two sentences with similar content can have a significant number of vectors. Thus, the relative distribution of terms may be identical in two sentences, but the absolute term frequency may be much more meaningful.

3) Jaccard Coefficient

Jaccard Coefficient is a statistical, computational method used to compare the similarity and variety of sample sets of sentences. Jaccard coefficients measure similarities between sets of samples described as the intersection size, divided from the size of the union of a collection of samples [20]. Between queries and dialogue sentences are calculated scores to classify the dialog, so we used the Jaccard Coefficient, if the set is closed, the higher the similarity of Jaccard. If A and B are empty, assume $J(A, B)$= 1

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A \cap B| + |A - B| + |B - A|} \quad (10)$$

Where, given a set A, the cardinality of A can be denoted as $|A|$ so calculate how many elements are in A, the intersection between two sets, namely A and B, can be denoted as $A \cap B$, and it shows all items on both sets. Whereas for the union between two sets, namely A and B, it is denoted as $A \cup B$ and shows all items set.

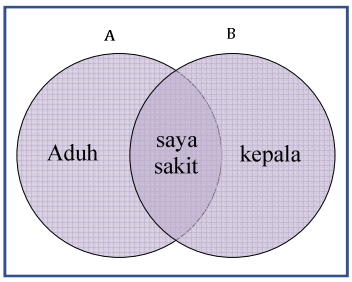

Fig. 5 Set Diagram of Jaccard Coefficient

Based on the Figure 5 shows a set diagram of sets A and B from a case example between the sentence dialog and the query. The important issue of the class with Jaccard Coefficient is to find a similar sentence from the dialog dataset that must be considered, namely the aspect of character similarity.

4) Finite State Machine

Finite State Machine (FSM) is used to model control and sequence processes in a system with a finite number of states. In particular, the actions of the system depending on the state do not depend only on the input to the system but also

on what happened earlier in the system. State machines are very important for determining systems with behaviors that depend on significant circumstances [21].

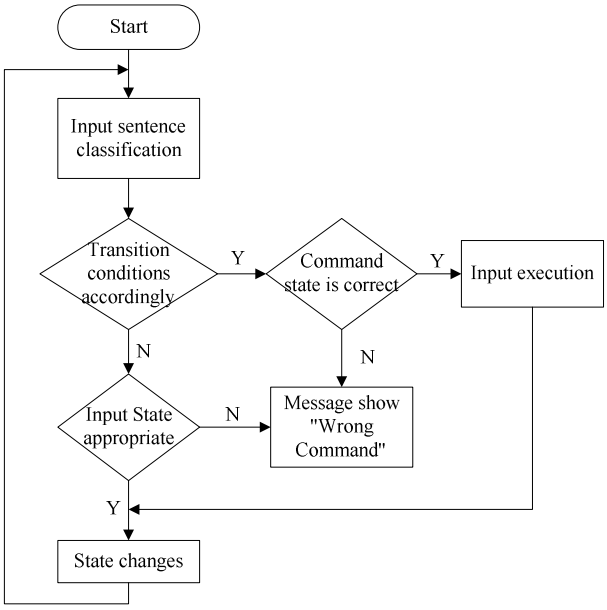

Fig. 6 Flowchart of Finite State Machine

Figure 6 is a flowchart for the sequential dialogue process, which depends on the conditions of the input state or event. If it is true, it executes the input, but if the state is wrong, then the transition condition will be visible. If it is true, then continue to run the state, and the state changes if all conditions above are wrong, then a message will appear in the command is not desired.

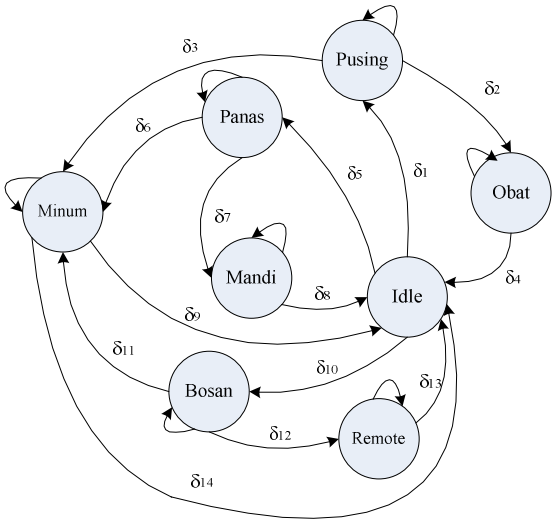

Fig. 7 Scenario Dialog Using Finite State Machine

The expected structure of the dialogue scenario is in Figure 7, where there are eight states with the S symbol that corresponds to the scenario dialogue. Transitions are marked with a symbol δ, which triggers changes between states. Dialogue management control based on the finite state machine is based on user requests. The system continues the current state, and the system responds based on the previous transition and circumstances [10].

In this study, it was designed for a management dialogue system with eight dialogue scenarios for robots that help the elderly in the home environment. Table I shows the eight scenarios in Indonesian.

TABLE I
SCENARIO DIALOG

Classification Dialog	Indonesian Scenario Dialog	Annotation
D1	Kepala saya sakit	My head hurts
D2	Saya ingin obat	I want medicine
D3	Hari ini cuacanya panas	Today the weather is hot
D4	Saya ingin mandi	I want to take a bath
D5	Sekarang saya bosan dirumah	Now I'm bored at home
D6	Saya ingin menonton televisi	I want to watch television
D7	Saya ingin minum	I want to drink
D8	Halo, robot	Hello, robot

Scenarios Dialogues are related to each other such as D1 related to D2 or D7, D3 is related to D4 or D7, while D5 is related to D6 and D7, which is the dialogue to help the elderly to carry out their activities at home assisted by robots.

5) Integrating Hardware and Software

The output of the dialogue management system with a comparison method between TF-IDF Cosine Similarity and Jaccard Coefficient and using Finite State Machine methods aim to give commands with natural language to the Bioloid Premium Robot. In Figure 8, the system structure for speech recognition is using Sensor Kinect as a microphone. The application of the method for management dialogues built-in Microsoft Visual Studio is using C # on a PC, then commands outputs are sent via serial communication between PCs and Arduino microcontrollers for Bioloid Robot controllers.

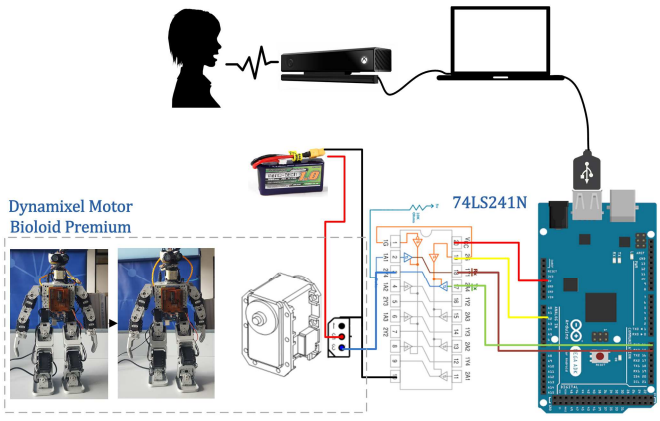

Fig. 8 Integrated System

The important thing for drive robots for communication for transmission data between Arduino and Dynamixel on Bioloid Premium Robot is using IC 74LS241N because communication between PC and Arduino is a full-duplex with 2 data lines. In contrast, dynamixel produces half-duplex serial communication, which is only one data line for communication. So the IC 74LS241N for serial data multiplexers can be used for communication with many dynamixel [2].

III. Results and Discussion

The environment for testing this data is carried out in the laboratory, which allows for the smallest possible noise disturbances. The location of the Kinect sensor in the static field with speech recognition distance remains with whom the person speaking. The dialogue system in management dialogue was tested starting from speech to text using the Google Cloud Speech API, so we tested for accuracy and response time of translation. Performed for eight dialogue scenarios for five times, then calculated the average value of the response time.

TABLE II
Response Time of Speech Recognition

Scenario	Accuracy (%)	Response Time (second)
D1	98	3.0275
D2	98	2.4
D3	100	2.86
D4	97.5	2.34
D5	100	2.7475
D6	100	2.75
D7	100	2.515
D8	95	2.32

Test results for time speech recognition using Google Cloud in table I show results that show if the average translation time range around 2.32 s until 3.0275 s. The long translation process is caused by the length of sentences that are transmitted to the cloud so that it takes longer because the process for speech recognition must go through the process of converting sound files into wav files and then transmitted into the cloud so that it turns into text. It depends on the speed of the internet. The average accuracy for speech recognition is 98.625%. The translation speed and accuracy of speech recognition also affects the management dialogue system process.

TABLE III
Tested People

Test Samples	Name	Ethnic Group	Gender	Age
I	Dinda Ayu P.	Jawa	Female	23
II	Hanif Fakhrurroja	Sunda	Male	38
III	Harika Ersya Putri	Melayu	Female	25
IV	Zakarias Januaji	Sunda	Male	25
V	Nova Agnes P.	Sunda	Female	20

Testing the dialogue for the management dialogue system was tested by five people of different gender, ethnicities, and ages. Because in Indonesia, it consists of many cultural tribes with different dialects. Furthermore, testing for this management dialogue is carried out for the eight dialogue scenarios that have been determined the same as the scenario, but in each classification, one dialogue scenario is carried out for two dialogues with different words. Each dialogue is tested ten times for 14 dialogues sentences with seven classifications of dialogue; everyone who tests says 570 words. Tests were conducted by comparing the accuracy of the classification of dialogue between cosine similarity with the Jaccard coefficient and testing the ability of FSM to control the flow of dialogue to make decisions.

TABLE IV
Dialog Management Accuracy

User Query	Accuracy (%)	
	TF-IDF Cosine Similarity	Jaccard Coefficient
Kepala saya sakit (My head hurts)	98	92
Kepala saya pusing sekali (My head is very dizzy)	92	84
Saya ingin obat (I want medicine)	98	94
Saya ingin obat sakit (I want sick medicine)	98	50
Hari ini cuacanya panas (Today the weather is hot)	100	100
Saya merasa panas nih (I feel hot)	100	100
Saya ingin mandi (I want to take a bath)	96	96
Saya ingin mandi agar tidak panas (I want to take a bath so it's not hot)	96	88
Sekarang saya bosan dirumah (Now I'm bored at home)	100	100
Bosan sekali saya dirumah (I'm so bored at home)	92	88
Saya ingin menonton televisi (I want to watch television)	100	100
Saya ingin menonton sinetron (I want to watch serial drama)	94	90
Saya ingin minum air (I want to drink water)	100	100
Sepertinya saya ingin minum saja (It seems like I want to drink it)	100	100

Evaluation results in Table IV show if the management dialogue system uses cosine similarity to classify sentences shows an average accuracy of 97.43%, while Jaccard similarity has an accuracy level of 91.57%. With the test case given, which uses several different words with the dialogue scene in the database, this indicates if the Cosine method can provide a way to process natural language. By changing the sentence structure, the Cosine similarity method has a high accuracy result for classifying sentences compared to the Jaccard Coefficient. Jaccard has a weakness if there are two classifications of similar sentences. This method does not have a reliable ability to classify based on the highest score. The Jaccard coefficient does not consider the frequency of terms. Cosine similarities can produce good accuracy.

The output from dialogue management, the robot does the commands given based on four commands by showing the location of "medicine", "drink", "remote", and "soap". Where the premium Bioloid Robot has 18 joints, while the one used for walking and shifting right and left is only the joint, which is represented by a red box in fig. 8.

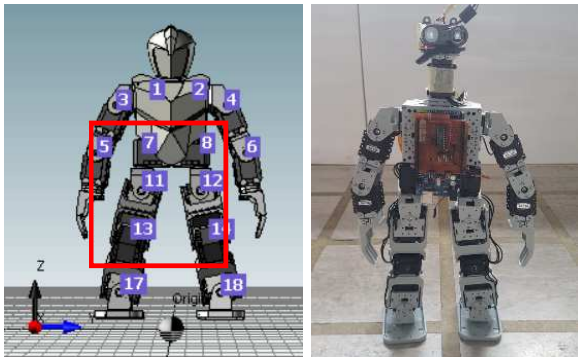

Fig. 9 (a) Display of Bioloid Robot with every Joint in Roboplus Software (b) Display of Robot Bioloid Implemented

Based on Figure 9, with the robot layout with the position of the object, the robot must walk to show the desired object command based on the output of the management dialogue. Path planning that has been determined to get to the coordinates of the object based on predetermined rules.

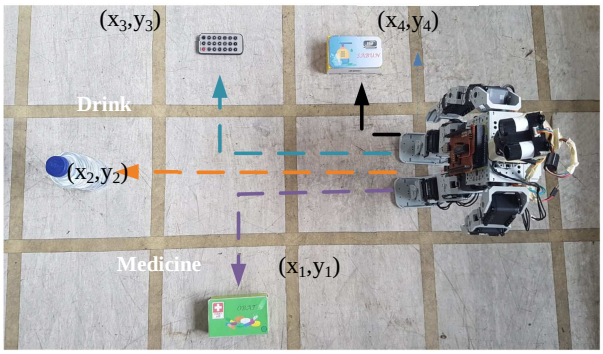

Fig. 10 Layout of Robot Against Object

The process of sampling changes in goal position from Bioloid Premium to show the location of this object is 28 steps with sampling time 0.03 s. Based on Figure 11, 12, and 13 show the change of servo goal position movement in the Bioloid Premium Robot, where this data is taken from the dynamixel library. Figure 11 shows the goal position of the dynamixel movement to walk based on the path to the object "medicine".

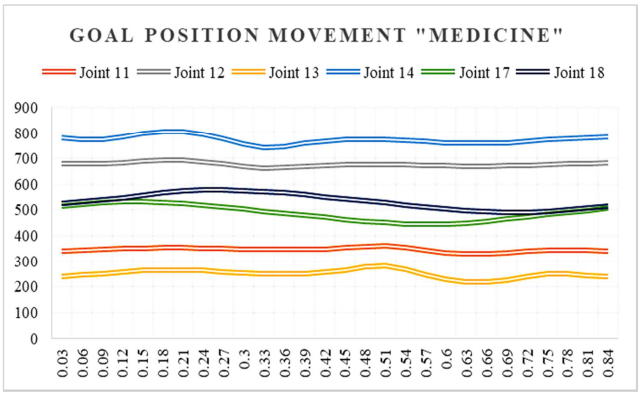

Fig. 11 Goal Position Movement for Command “Show Medicine”

In addition, in figure 12 shows the movement of the goal position from the joint to show the movement towards the object "medicine".

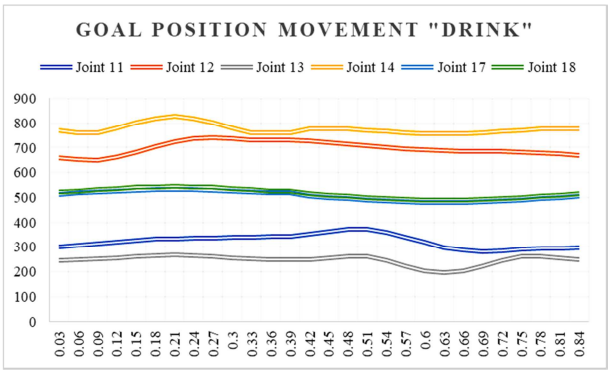

Fig. 12 Goal Position Movement for Command “Show Drink”

Figure 13 shows the goal position on the bioloid collaborative robot shows the movement to walk towards "soap" and "remote" objects because of the same path with the same movement.

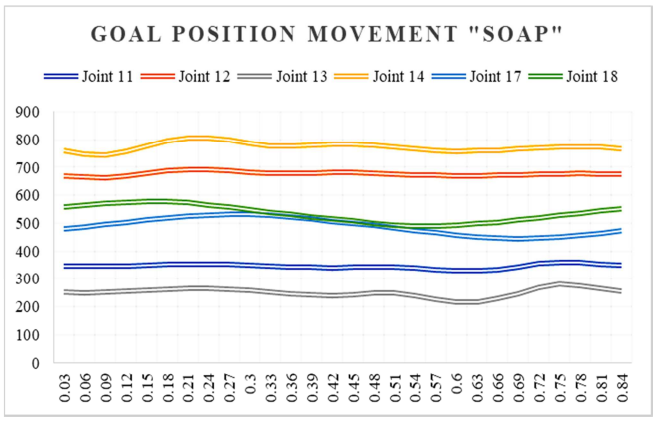

Fig. 13 Goal Position Movement for Command “Show Soap”

When there is an input form command from the management dialogue to show four objects, the PC sends serial data to the Arduino microcontroller. The servo joint on the Bioloid moves according to the instructions.

IV. Conclusion

The management dialogue proposed in our study is using the TF-IDF algorithm to give the weight of the term for its relationship with the dialogue sentence. Sentence classification is done by cosine similarity method with calculating the highest score from the highest cosine angle value and finite state machine to set the sequence flow dialogue. Results for time speech recognition using Google Cloud Speech as an engine for speech to text show that average translation time is fastest at 2.4 s. In contrast, at the longest time in 2.75 s with average accuracy 98.625%, it because of the extended effect of sentences sent to the cloud. Tests for management dialogue were carried out on five people with 14 dialogues conducted ten times with 570 words for each person, comparing the accuracy of the classification of dialogue between cosine similarity with the Jaccard coefficient. Experimental result for dialogue management using Cosine Similarity with an average accuracy of 97.43 %. Moreover, if using a Jaccard coefficient shows an average accuracy of 91.57%.

Evaluation results are seen if the cosine similarity method compared with the Jaccard coefficient results in high accuracy with the cosine similarity method. Because the output value of the cosine angle is closer to the target and higher than the Jaccard, and again cosine uses an algorithm to consider the frequency of terms. At the same time, Jaccard only calculates how many words are contained. From these results, it can be concluded TF-IDF Cosine Similarity method and using the finite state machine can provide good results for sentence classification and answer retrieval and run the appropriate dialogue scenario. This management dialogue can control the robot to show the location of the object.

ACKNOWLEDGMENT

This work was supported by the Post Graduate Team Research 2018 from the Ministry of Research, Technology and Higher Education, Republic of Indonesia.

REFERENCES

[1] T. H. Bui, "Multimodal Dialogue Management - State of the art," 2006.

[2] D. A. Maharani, H. Fakhrurroja, Riyanto, and C. Machbub, "Hand gesture recognition using K-means clustering and Support Vector Machine," in *ISCAIE 2018 - 2018 IEEE Symposium on Computer Applications and Industrial Electronics*, 2018, pp. 1–6.

[3] L. Meng and M. Huang, "Dialogue intent classification with long short-term memory networks," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 10619 LNAI, pp. 42–50, 2018.

[4] M. C. Hsieh, W. S. Hung, S. W. Lin, and C. H. Luo, "Designing an assistive dialog agent for a case of spinal cord injury," in *Proceedings - 2009 9th International Conference on Hybrid Intelligent Systems, HIS 2009*, 2009.

[5] K. Sadohara *et al.*, "Sub-lexical Dialogue Act Classification in a Spoken Dialogue System Support for the Elderly with Cognitive Disabilities," *Proc. Fourth Work. Speech Lang. Process. Assist. Technol.*, pp. 93–98, 2013.

[6] S. Schwarzler, J. Schenk, G. Ruske, and F. Wallhoff, "A multi-agent framework for a hybrid dialog management system," in *2009 IEEE International Conference on Multimedia and Expo*, 2009, pp. 958–961.

[7] H. Holzapfel, "A dialogue manager for multimodal human-robot interaction and learning of a humanoid robot," *Ind. Rob.*, vol. 35, no. 6, pp. 528–535, 2008.

[8] C. Lee, Y. S. Cha, and T. Y. Kuc, "Implementation of dialogue system for intelligent service robots," in *2008 International Conference on Control, Automation and Systems, ICCAS 2008*, 2008.

[9] A. Raux and M. Eskenazi, "A Finite-State Turn-Taking Model for Spoken Dialog Systems," in *Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the ACL*, 2009.

[10] S. Yi and K. Jung, "A Chatbot by Combining Finite State Machine, Information Retrieval, and Bot-Initiative Strategy," *Alexa Price Proc.*, pp. 1–10, 2017.

[11] B. Su, T. Kuan, S. Tseng, J. Wang, and P. Su, "Improved TF-IDF weight method based on sentence similarity for spoken dialogue system," in *2016 International Conference on Orange Technologies (ICOT)*, 2016, pp. 36–39.

[12] C. Lee, S. Jung, S. Kim, and G. G. Lee, "Example-based dialog modeling for practical multi-domain dialog system," *Speech Commun.*, 2009.

[13] H. Fakhrurroja, D. A. Permatasari, A. Purwarianti, and C. Machbub, "Dialogue Management for Human Robot Interaction Using Artificial Intelligence Markup Language," in *ICEECS 2018 International Conference on Electrical Engineering and Computer Science*, 2018.

[14] X. Zhang and Y. LeCun, "Character-level Convolutional Networks for Text Classificatio," in *Advances in Neural Information Processing Systems 28*, 2015.

[15] D. Petcu, C. Craciun, and M. Rak, "Towards a Cross Platform Cloud API - Components for Cloud Federation," in *CLOSER*, 2011.

[16] Google, "Google Speech API," *Google Cloud Platform*, 2017.

[17] M. Assefi, G. Liu, M. P. Wittie, and C. Izurieta, "An Experimental Evaluation of Apple Siri and Google Speech Recognition," *Proccedings 2015 ISCA SEDE*, 2015.

[18] C. C. Chiu *et al.*, "State-of-the-Art Speech Recognition with Sequence-to-Sequence Models," in *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, 2018.

[19] C. D. Manning, P. Raghavan, and H. Schutze, *An Introduction to Information Retrieval*. Cambridge University Press, 2009.

[20] N. Agarwal, M. Rawat, and M. Vijay, "Comparative Analysis Of Jaccard Coefficient and Cosine Similarity for Web Document Similarity Measure," *Int. J. Adv. Res. Eng. Technol.*, 2014.

[21] H. Gomaa, *Real-Time Software Design For Embedded Systems*. New York: Cambridge University Press, 2016.