

SeerahBot: An Arabic Chatbot About Prophet's Biography

Soumeya Mahamat Yassin¹, Mohammad Zubair Khan²

^{1,2}Department of Computer Science, College of Computer Science and Engineering, Taibah University,
Madinah, Saudi Arabia

Correspondence should be addressed to Soumeya Mahamat Yassin; somayh88@gmail.com

Copyright © 2021 Made Soumeya Mahamat Yassin et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

ABSTRACT- Chatbots have become very popular in our times and are used in several fields. The emergence of chatbots has created a new way of communicating between human and computer interaction. In this study, we have developed a Chatbot in Arabic language, which is a program that can understand and analyze Arabic language. The Chatbot developed in this study contributes to its users about the Prophet's biography and provide them with valuable information. SeerahBot will be the first chatbot dedicated to the Prophet's biography. We implemented this system by using two methods first build retrieval based chatbot from scratch using natural language processing NLP and machine learning techniques and second technique we used a chatbot design platform. To facilitate access to SeerahBot, we have made it available on Telegram. SeerahBot was tested, and the results were acceptable.

KEYWORDS- Artificial intelligence; NLP; Arabic chatbot

I. INTRODUCTION

Artificial intelligence (AI) has had a significant effect on our everyday lives in the digital era by developing innovative software and devices known as intelligent agents that can perform a variety of functions. Chatbot is AI program which played an important role in the Human-Computer Interaction (HCI) field, which allows communication between humans and computers by using natural language [1]. A chatbot, as defined by the dictionary, is a computer program designed to simulate human conversation, usually over the Internet [2]. In the scientific literature, Chatbot was considered as a conversational agent (CA), which has ability to talk with user about a specific subject by using the natural language processing (NLP). A Chatbot, also called a "Chatter Robot," is a software application that mimics human conversations in its natural format, which contains textual material as well as oral communication with the help of AI techniques. There are two kinds of chatbots: rule-based and smart machine-based. Chatbots can be classified into different categories such as e-commerce (via chat), education, food, entertainment, games, medical (Heath), Human Resource (HR), news, travel, etc. We suggest in this research a new Arabic Chabot called SeerahBot. The Chatbot developed in this study contributes to its users about the Prophet's biography and provide them with valuable information and answer their enquiries. SeerahBot will be the first chatbot

dedicated to the Prophet's biography. We implemented this system by using two methods first build chatbot from scratch using NLP and machine learning techniques and second technique we used a chatbot design platform. To facilitate access to SeerahBot, we have made it available on Telegram. To assess SeerahBot, a usability test was conducted. The remainder of this paper is organized as follows. In section 2, we take a look at the history of chatbots from their inception to the present day. In section 3, we describe some of related work. In section 4, we present our chatbot creation methodology. Finally, we present the results of the SeerahBot evaluation.

II. CHATBOT BACKGROUND

In 1950, was published an article "Computing Machinery and Intelligence" written by Alan Turing, He was wondering whether a computer program could converse with a group of people without perceiving that their interlocutor was artificial. A first-ever bot ELIZA was constructed in 1966 [3]. Figure 1 shows the most popular and famous ELIZA conversation, in which ELIZA simulated a psychotherapist's operation giving very sophisticated therapeutic advice [4]. ELIZA uses pattern matching technique to categorize the user input as <pattern> and then give an appropriate replay kept in < template> [5]. Its communication and information abilities were limited, but it served as a model for the development of other chatbots. In 1995, the term "Artificial Linguistic Internet Computer Entity" was coined. A.L.I.C.E. or Alicebot was the first online chatbot inspired by ELIZA [6]. It was an NLP bot. She converses with a human by using heuristical pattern matching rules on the human's input. A.L.I.C.E was developed using a new language called Artificial Intelligence Markup Language (AIML), which was created specifically for this purpose, therefore this is the main difference between A.L.I.C.E and ELIZA [7].

```

Men are all alike.
IN WHAT WAY
They're always bugging us about something or other.
CAN YOU THINK OF A SPECIFIC EXAMPLE
Well my boyfriend made me come here.
YOUR BOYFRIEND MADE YOU COME HERE
He says I'm depressed much of the time.
I AM SORRY TO HEAR YOU ARE DEPRESSED
::
WHO ELSE IN YOUR FAMILY TAKES CARE OF YOU?
My father
YOUR FATHER
You are like my father in some ways.
WHAT RESEMBLANCE DO YOU SEE
You are not very aggressive but I think you don't want me to notice that.
WHAT MAKES YOU THINK I AM NOT AGGRESSIVE
You don't argue with me.
WHY DO YOU THINK I DON'T ARGUE WITH YOU
You are afraid of me.
DOES IT PLEASE YOU TO BELIEVE I'M AFRAID OF YOU
My father is afraid of everybody.
WHAT ELSE COMES TO MIND WHEN YOU THINK OF YOUR FATHER
Bullies.
DOES THAT HAVE ANYTHING TO DO WITH THE FACT THAT YOUR
BOYFRIEND MADE YOU COME HERE

```

Fig. 1: Sample ELIZA dialogue from Weizenbaum (1966)

Siri was created by Apple in 2010 as a voice-controlled personal assistant, and it has been available for many years. Siri is intended to provide you with a streamlined way of communicating with your iPhone, iPad, and other devices. Users can make conversations with Siri by using voice commands, and it includes integration with audio, video, and image files [8]. Siri uses a set of internet resources to make suggestions and respond to user requests. With continued use, it adapts to users' language usages, searches, and desires. Despite the fact that Siri is a sophisticated and advanced device, it has several flaws. It necessitates the use of the internet. It is multilingual, but only supports a few languages and needs an internet connection [9]. As shown in Figure 2 according to Scopus [8,] there was a rapid increase interest in the chatbots field after 2016.

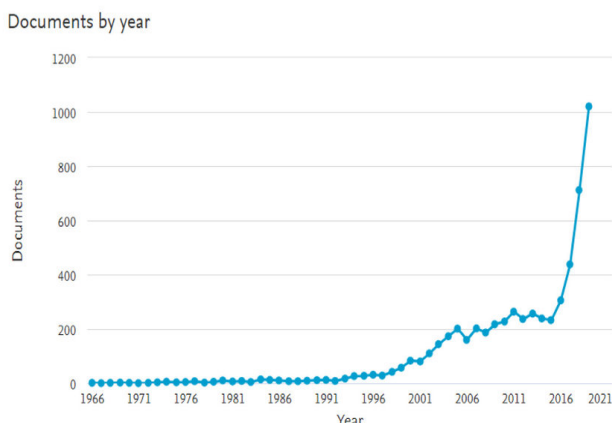


Fig 2: Search results in Scopus by year from 2000 to 2019 for “chatbot” or “conversation agent” as keywords.

III. RELATED WORK

CAs have developed in several languages and fields. In this section, we will do survey on English and Arabic CAs. This paper [10], describes a medical chatbot that can be used to replace traditional disease diagnosis and treatment referral methods using machine learning. A chatbot can serve as a doctor. The chatbot functions as a user program. The user of

this application will specify their symptoms to the chatbot, and the chatbot will then specify the health steps to be taken. Since the dataset contains general information regarding symptoms and diseases, the chatbot instance will provide the user with disease and treatment information. After observing the symptoms of the various patients, it eventually predicts the disease to the patient and offers a link where information about the treatment is available. In 2020 Naous, T., Hokayem, C., & Hajj, H [11], were proposed the first model for Arabic empathetic conversational bots and a dataset of conversations in Arabic. The proposed model is a sequence-to-sequence (Seq2Seq) model with LSTM units combined with Attention. The dataset is available in the Arabic language, they translated the EmpatheticDialogues dataset [12], which is the only available and include dataset in English for building empathetic chatbots. They used the Googletrans API [13] to perform the translations from English to Arabic. To assess the quality of the translation they chose 100 random translated samples and compared them with the original English samples. The results indicated that only 6 of the 100 randomly chosen samples were found unreasonable while the rest of the samples reasonable. These types of errors are rare in the generated conversation dataset and the translation system was showed an accuracy of 96% on a sample of the data, which was deemed as sufficient for the purpose of model development and training conversational bots. The proposed model is evaluated by using the Perplexity (PPL) automated metric and the BLEU score as an additional metric for evaluation, with an embedding dimension of 500 reached state of the art performance for Arabic with a PPL of 38.6 and a BLEU score of 0.5. human ratings are an important part of the overall evaluation. They collected ratings from 50 speakers of the Arabic language and rate them in terms of Empathy, Relevance, and Fluency, by answering some questions. Human evaluation of the generated responses also validated the success of the proposed model, shows performance for Arabic with average levels of Empathy and Fluency reaching of 3.7 and 3.92 respectively but the Relevance metric was at 3.16 the model did not always stay on topic while responses sometimes go off topic. In 2020 Al-Ghadhban, D., & Al-Twairesh, N [14] proposed Saudi dialect chatbot called Nabiha. This bot serves the student in information technology IT department in King Saud University by doing conversation with bot in Saudi dialect. The goal of this chatbot is communicate with the students and answers their questions about the courses offered in the IT department or any query related to their academic progress. Nabiha developed by using pattern matching and AIML.

IV. METHODOLOGY

In this section we design and implementation of the proposed Chatbot.

A. Implementation of chatbot model from scratch approach

a. System Architectural Design

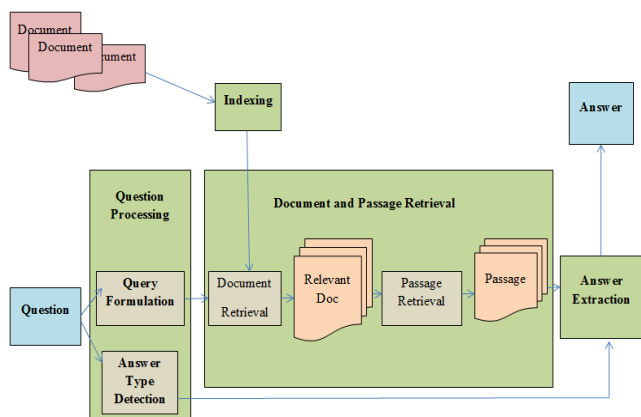


Fig 8: Question Answering (QA) Pipeline

b. Dataset Collection and Preparation

We created our own dataset by forming 200 questions and answers about the Prophet's biography by using Seerah book (اللولؤ المكنون في سيرة النبي المأمون) [15] which is related to our topic in order to obtain correct and accurate information. Then we arranged and placed the data in a text file.

c. Question Processing

The main goal of the question-processing phase is to extract the query, the keywords passed to the IR system to match potential documents. we summarize the two most commonly used tasks, query formulation and answer type detection as follow:

• Query Formulation

Query formulation is the task of creating a query—a list of tokens—to send to an information retrieval system (chatbot) to retrieve document that might contain answer strings.

• Answer Type Detection

Chatbot will check the existence of the most sentence of the text which is similar to the user queries. Queries can be categorized as the following types that are illustrated in Table 1 and each type is paired to a distinct type of an answer.

Table 1: Q & A Type

Answer Type	Query Type (Question)
معلومات تخص غير العاقل	ما
معلومات تخص العاقل	من
الطريقة	كيف
الزمان	متى – كم
المكان	أين

d. Document and Passage Retrieval

Chatbot will find the most sentence of the text, which was already processed by Natural Language Processing, is similar to the user queries. Then, Chatbot will get this sentence in order to return it as a response back to the user.

e. Answer Extraction

The final stage of question answering is to extract a specific answer from the passage.

B. System Design and Implementation

We have implemented the chatbot code using Colaboratory, or in short “Colab”, is a product from Google Research. It allows anyone to write and execute python code through the browser. More technically, Colab is a free-hosted Jupyter notebook environment that runs completely in the cloud. The most important aspect of this tool does not require a setup to use. In addition, it supports writing and executing code in Python.

The codes start preparation by the following steps. Each step is implemented and written in Appendixes.

1. Installing the package NLTK. See Appendix A. -NLTK stands for Natural Language Tool Kit which has platform to write python programs in order to work with some data of human natural language.
2. Importing all the required libraries. See Appendix B.
3. Load the dataset and convert every text into lowercase. See Appendix C.
4. Data pre-processing. See Appendix D. Data cleaning and preprocessing by tokenizing the text into a list of sentences and word.
5. Creating dictionary to remove punctuations. -Creating a function to return a list of lemmatized lower-case words from the tokenized text after the punctuations are removed. See Appendix E.
6. Creating a greeting for user’s input and Chatbot response as well as returning the greeting response randomly. See Appendix F.
7. Generating the response. See Appendix G.

In this step, machine learning aspect of Chatbot is taking a place as the following:

1. It takes user response (query).
2. It appends user response to the end of the sentence list of the tokenized text.
3. It does the product of term frequency (TF) and inverse document frequency (IDF). -Term frequency measures how frequently a term occurs in a document. -Inverse document frequency measures how rare a term is.
4. It converts the tokenized text to a matrix of TF-IDF.
5. It returns the index of the most similar sentence to the user response.
6. It prints the most similar sentence.
7. If there is no similar sentence to the user response, it prints an apologize message.
8. Lastly, creating a loop for the conversation between Chatbot and the user. See Appendix H.

C. Experimental Results

In this section we show the result of our Chatbot program which is called EduBot.

It produces four observable functionalities. These functionalities are mentioned as the following:

1. Response back to user’s greeting.
2. Apologize to user if any misconception has occurred.

3. End the conversation with the user by displaying ending response after the user types his/her ending message like: (thank you شكرًا or bye إلى اللقاء).
4. Return a precise answer to the user about his/her query.

Figure 9 below shows the result of the first function.

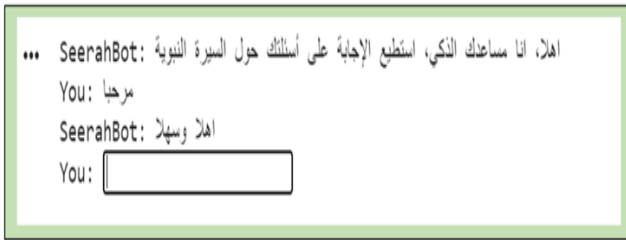


Fig 9: Response back to user's greeting

Figure 10 below shows the result of the second function.

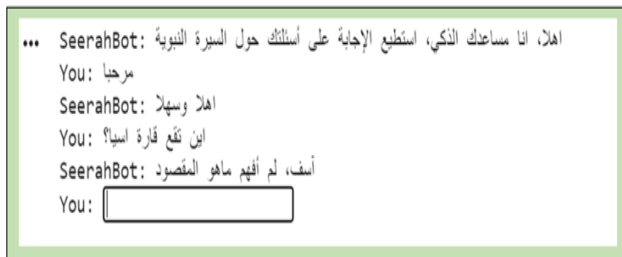


Fig 10: Apologize to user if any misconception has occurred

Figure 11 below shows the result of the third function.

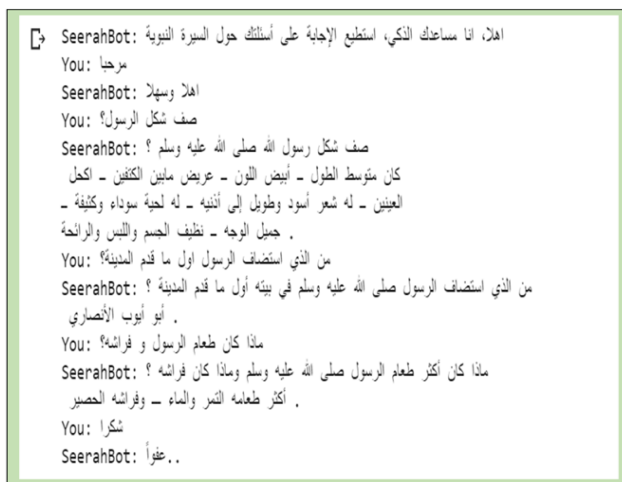


Fig 11: End the conversation with the user by displaying ending response after the user types his/her ending message

Figure 12 below shows the result of the fourth function which is.

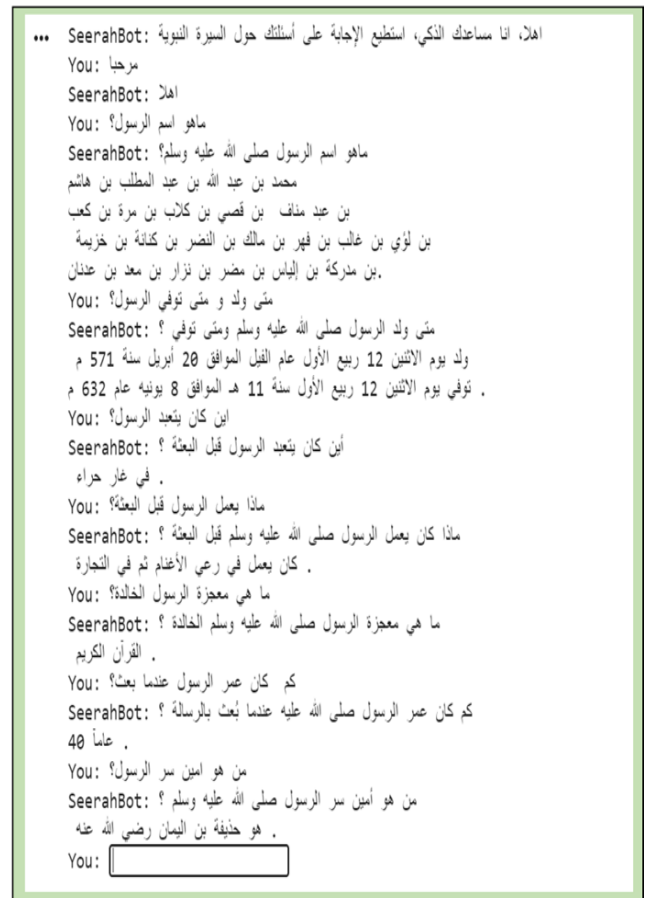


Fig 12: Return a precise answer whether short or long answer to the user about his/her query

D. Execution of chatbot model using platform

In this section we first, build the chatbot model by using Microsoft Azure platform and then test it in the Telegram. Microsoft Azure, also known as Azure, is a cloud computing service developed by Microsoft that allows users to build, test, deploy, and manage applications and services through Microsoft-managed data centres [16].

a. Building the chatbot model using Microsoft Azure platform

After we created account on Microsoft Azure. We created our knowledge base and published it as shown in Figure 13, after that Create our bot as shown in Figure 14.

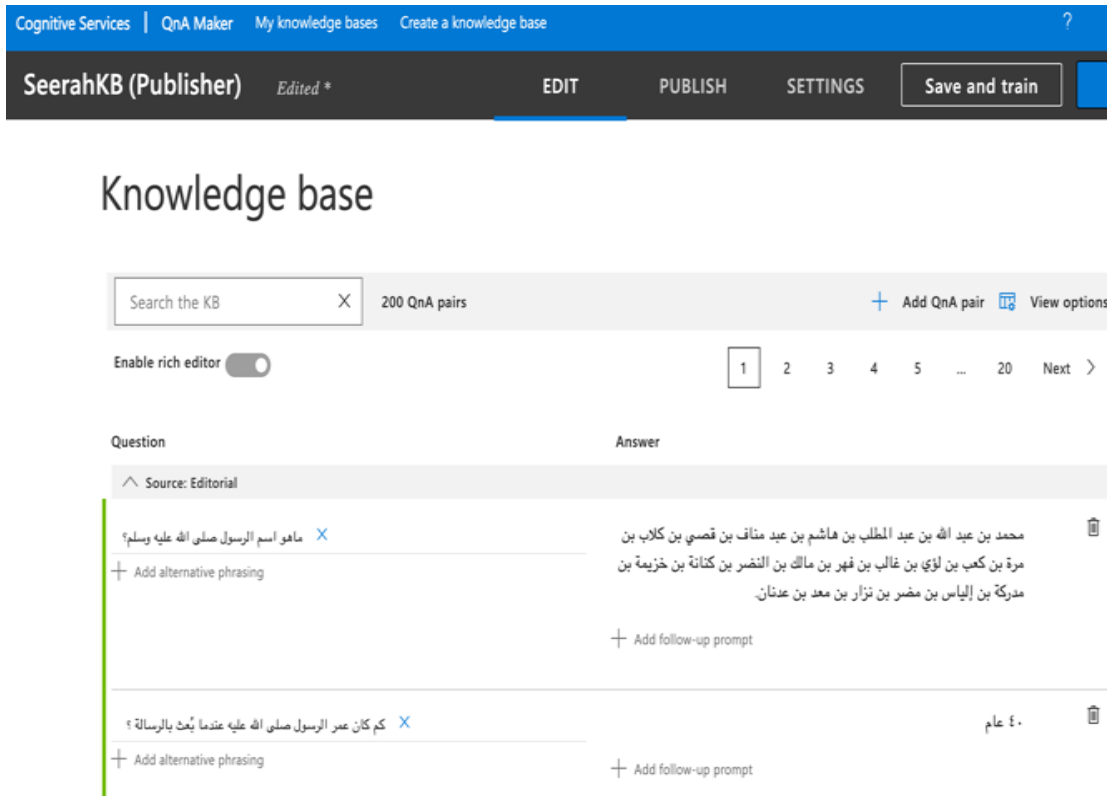


Fig. 13: Create KB

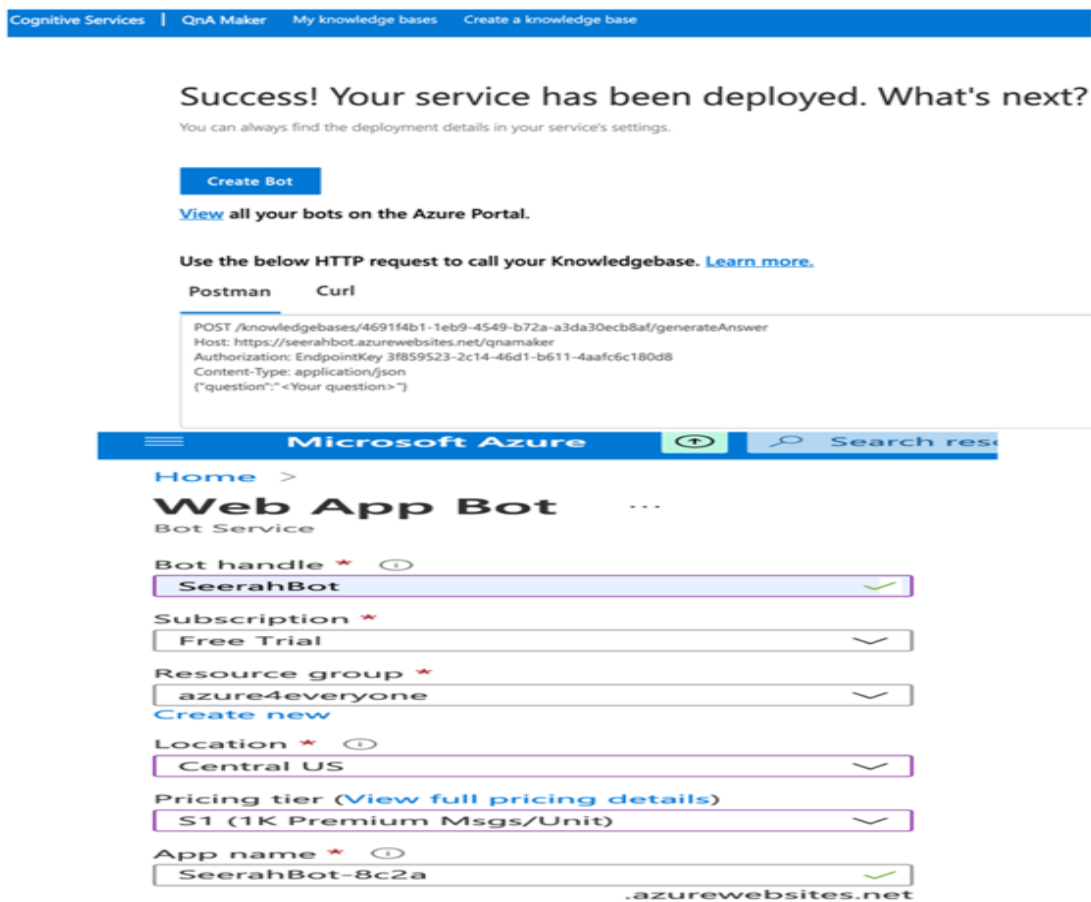


Fig. 14: Create Chatbot

Before we connect our bot to Telegram channel, we are going to explain the steps of creating telegram bot:

Step 1. Enter Telegram messenger and choose @Botfather in the search tab.

Step 2. We Request from @Botfather a new bot. Then we choose name for our bot as shown in the Figure 15.

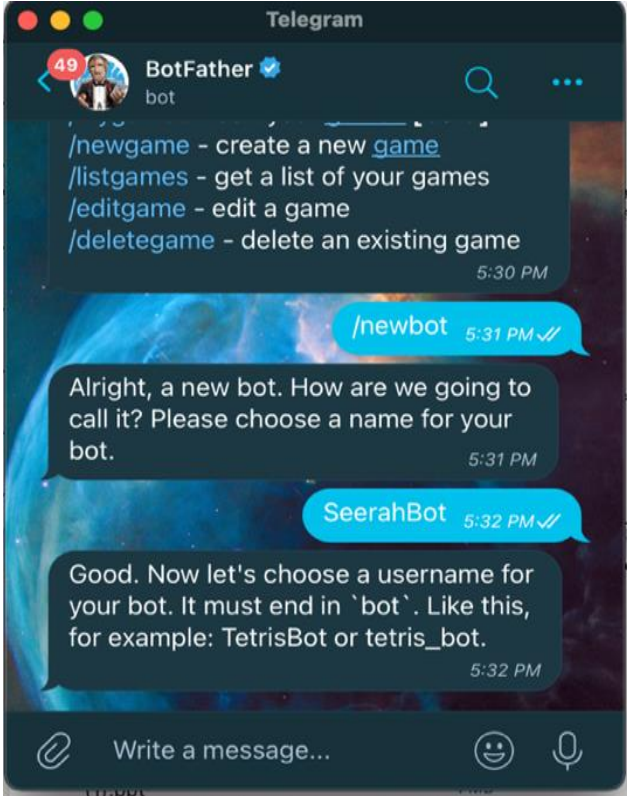


Fig 15: Create chatbot in Telegram & choose the name of our bot

Step 3. Copy the Telegram bot's access token as shown in Figure 16.

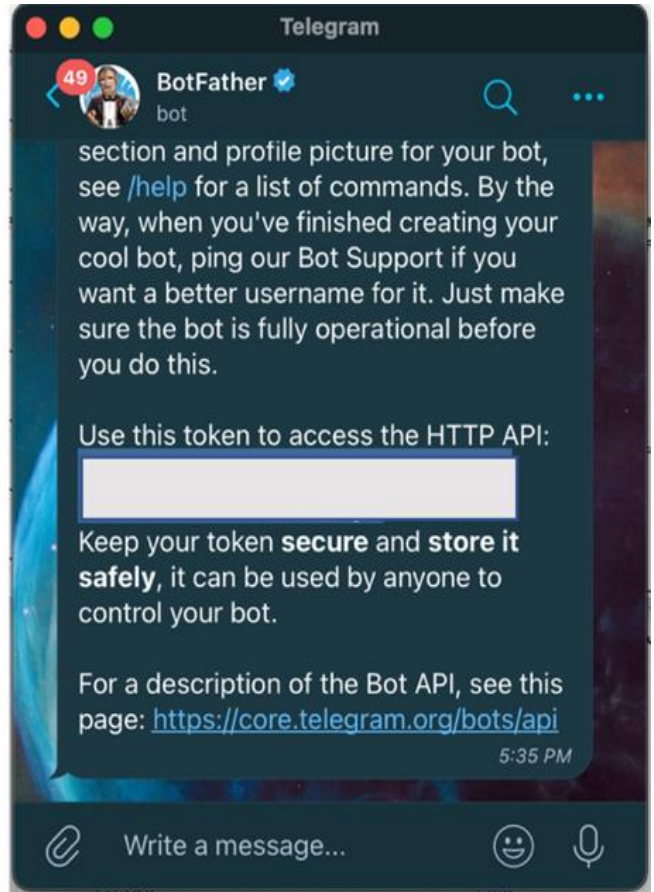


Fig 16: Chatbot token access

After we created Telegram bot's access token, we go to our bot's channels section in the azure portal and click on telegram as shown in Figure 17 chatbot token access.

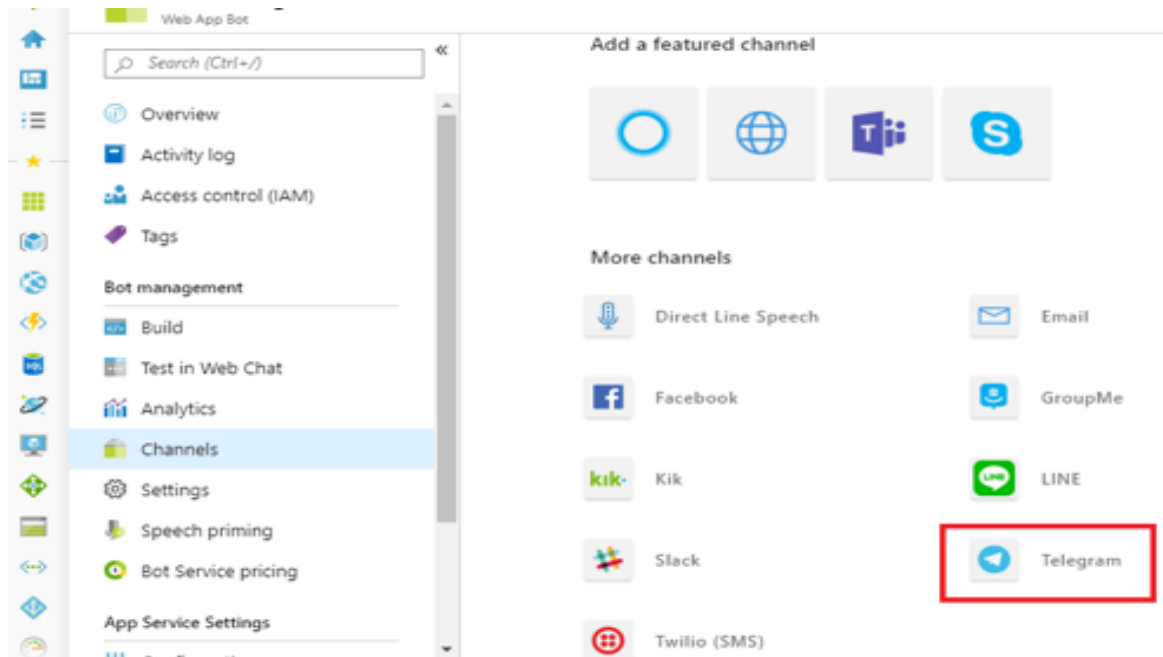


Fig 17: Choose telegram channel

Paste the token we copied into the Access Token field as shown in the figure Figure 18.

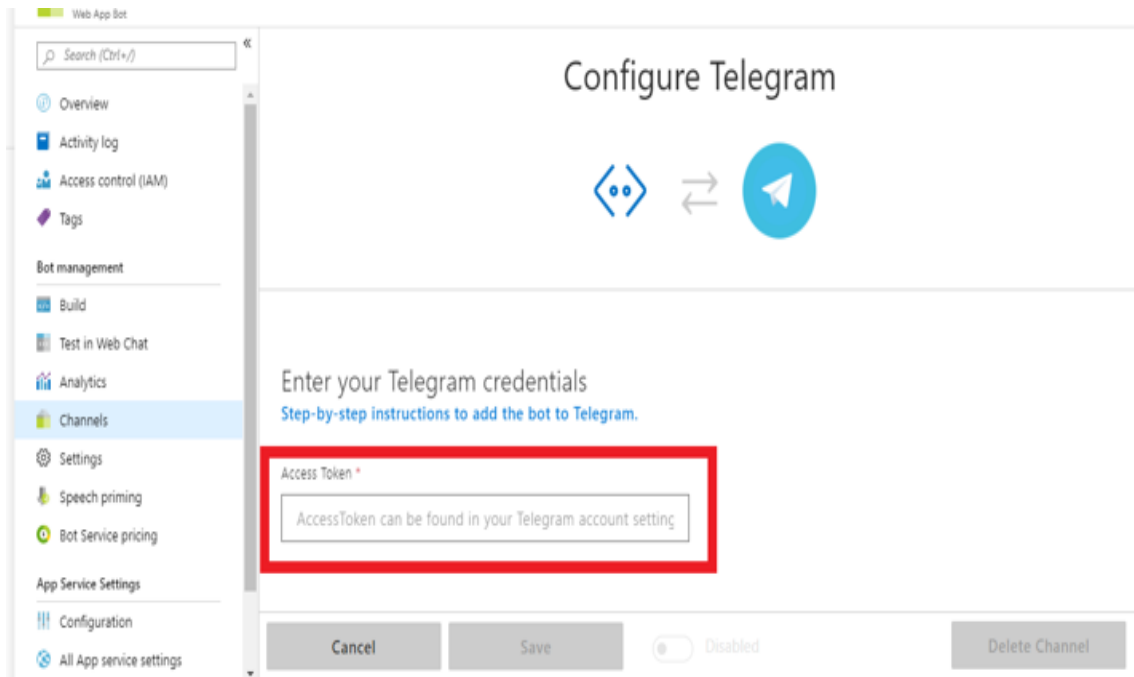


Fig. 18: Paste Access Token

our bot is now successfully configured as shown in figure Figure 19 to communicate with users in Telegram.

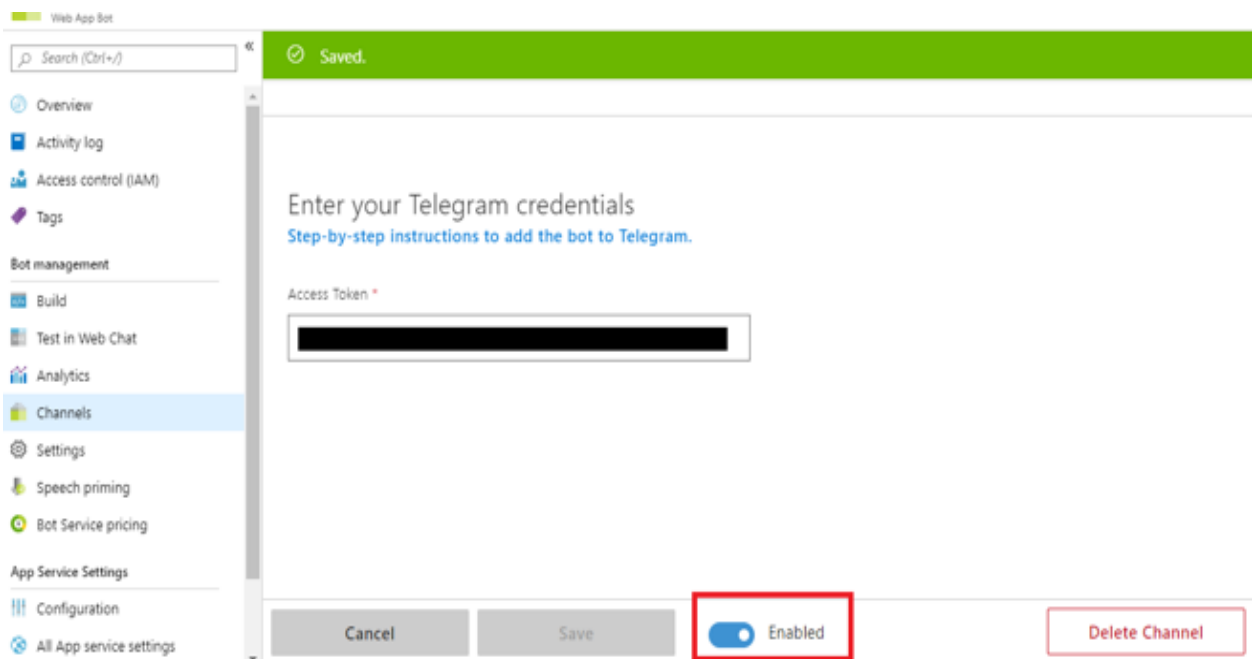


Fig 19: Configured bot to Telegram

b. Executing the chatbot in Telegram



Fig 20: Conversation with SeerahBot on Telegram

c. Evaluation

Radziwill and Benton [17] proposed a methodology to assess the quality of chatbots; they identified several quality issues and attributes that aid in the quality assessment of chatbots based on ISO 9241. We evaluated SeerahBot using these attributes [20]. ISO 9241 is a set of international standards for HCI and usability [18]. ISO 9241 defines usability as “Software is functional when it enables the user to perform his task effectively, efficiently, and satisfactorily in the specified context of use.” [18]. Thus, the three most important characteristics are effectiveness, efficiency, and satisfaction. Effectiveness refers to how well the system achieves the user's expectations, i.e., quality and completeness. Efficiency refers to how well resources are used to meet the user's goals. Satisfaction refers to customer satisfaction: how they feel about using the program. To ensure the usability of the SeerahBot, we asked 14 person to use it and provide feedback through a questionnaire. We attempted to ask questions in the questionnaire that reflected effectiveness, efficiency, and satisfaction categories, as well as quality attributes. We concentrated on performance to assess effectiveness. There are several quality attributes related to it such as robustness to unexpected input. To assess efficiency, we focused on functionality and humanity, as well as quality attributes such as linguistic accuracy of outputs and the ability to react to specific questions. To assess satisfaction, we focused on affect, ethics, behaviour, and accessibility, as well as quality attributes such as greetings. Thus, we divided our questions into three several as shown in Table 2.

d. Results

We obtained the results in Table 2 after checking the usability of our chatbot in various categories. During the usability test with students, SeerahBot was unable to reply to some questions. Because not their question not existed in SeerahBot's knowledge base. To summarize this section, we can conclude that the results of the first experiment with the SeerahBot were somewhat appropriate after performing

usability testing of the SeerahBot and based on the questionnaire results. However, SeerahBot still needs to be enhanced by extending the dataset.

Table 2: Usability Testing Results

Category	Question	Results			
The methodology of SeerahBot in conducting conversation.	How was the experience of your communication with SeerahBot?	Easy 14/14 (100%)		Difficult —	
	Have you been welcomed?	Yes 14/ 14 (100%)		No —	
Quality of information provided by SeerahBot	How would you evaluate the level of information you got from SeerahBot?	Bad —	Acceptable 3/14 (21.43%)	Good 4/14 (28.57%)	Excellent 7/14 (50%)
	Was SeerahBot accurate in giving answers?	Not accurate 2/14 (14.29%)	Somewhat accurate 7/14 (50%)	Very precise 5/14 (35.71%)	
SeerahBot's abilities.	Did SeerahBot interact with your questions?	Yes 14/14 (100%)		No —	
	Has SeerahBot replied promptly?	Slow 1/14 (7.14%)		Fast 13/14 (92.86%)	
Conversation with SeerahBot in general.	How satisfied are you with SeerahBot?	Dissatisfied —		Somewhat satisfied 4/14 (28.57%)	Totally satisfied 10/14 (71.43%)
	Will you chat and ask SeerahBot again?	Yes 13/14 (92.86%)		No 1/14 (7.14%)	

V. CONCLUSION

In this study, we have presented a new Arabic Chabot called SeerahBot. The Chatbot developed in this study contributes to its users about the Prophet's biography and provide them with valuable information. We implemented this system by using two methods first build chatbot from scratch using NLP and machine learning techniques and second technique we used a chatbot design platform. To facilitate access to SeerahBot, we have made it available on Telegram. SeerahBot was tested by some users, and the results were acceptable. However, the limitation of SeerahBot: The dataset should be expanded. As a result, we intend to implement this improvement in the future. Furthermore, we want to make SeerahBot available on other channels such as WhatsApp and FaceBook.

REFERENCES

- [1] Shawar, B.A., Atwell, E.: Different measurements metrics to evaluate a chatbot system. In: Proceedings of the Workshop on Bridging the Gap: Academic and Industrial Research in Dialog Technologies, Association for Computational Linguistics, pp. 89–96 (2007)
- [2] Chatbot | definition of chatbot in english by Lexico Dictionaries, . (2019). Lexico Dictionaries | English website: <https://www.lexico.com/en/definition/chatbot>. (Retrieved 16 July 2019).
- [3] Weizenbaum, J.: ELIZA—a computer program for the study of natural language communication between man and machine. *Commun. ACM* 9(1), 36–45 (1966)
- [4] Daniel Jurafsky, James H.(2019). *Martin.Speech and Language Processing*. Third Edition

- [5] Brandtzaeg, P. B., & Følstad, A. (2017, November). Why people use chatbots. In International Conference on Internet Science (pp. 377-392). Springer, Cham.
- [6] Wallace, R.S.: The anatomy of ALICE. In: Parsing the Turing Test, pp. 181–210. Springer, Dordrecht (2009)
- [7] https://en.wikipedia.org/wiki/Artificial_Linguistic_Internet_Computer_Entity
- [8] Adamopoulos, E., & Moussiades, L. (2020). Chatbots: History, technology, and applications. Machine Learning with Applications, 2, 100006.
- [9] <https://en.wikipedia.org/w/index.php?title=Siri&oldid=942524254>
- [10] Mathew, R. B., Varghese, S., Joy, S. E., & Alex, S. S. (2019, April). Chatbot for disease prediction and treatment recommendation using machine learning. In 2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI) (pp. 851-856). IEEE.
- [11] Naous, T., Hokayem, C., & Hajj, H. (2020, December). Empathy-driven Arabic Conversational Chatbot. In Proceedings of the Fifth Arabic Natural Language Processing Workshop (pp. 58-68).
- [12] Rashkin, H., Smith, E. M., Li, M., & Boureau, Y. L. (2018). Towards empathetic open-domain conversation models: A new benchmark and dataset. arXiv preprint arXiv:1811.00207
- [13] <https://pypi.org/project/googletrans/>
- [14] Al-Ghadhban, D., & Al-Twaresh, N. (2020). Nabihah: An Arabic dialect chatbot. Int. J. Adv. Comput. Sci. Appl, 11(3), 1-8.
- [15] العازمي، موسى. (٢٠١٣). اللؤلؤ المكنون في سيرة النبي المأمون
- [16] https://en.wikipedia.org/wiki/Microsoft_Azure
- [17] N. M. Radziwill and M. C. Benton, "Evaluating Quality of Chatbots and Intelligent Conversational Agents," ArXiv Prepr. ArXiv170404579, 2017.
- [18] A. Abran, A. Khelifi, W. Suryan, and A. Seffah, "Usability meanings and interpretations in ISO standards," Softw. Qual. J., vol. 11, no. 4, pp. 325–338, 2003.

APPENDIX

Appendix A

```
pip install nltk
```

Appendix B

```
import nltk
import numpy as np
import random
import string
import warnings
warnings.filterwarnings('ignore')
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.metrics.pairwise import cosine_similarity
nltk.download('punkt')
nltk.download('wordnet')
```

Appendix C

```
f=open('ask.txt','r',errors = 'ignore')
raw=f.read()
raw=raw.lower()
```

Appendix D

```
word_tokens = nltk.word_tokenize(raw)# converts to list of words
sent_tokens = nltk.sent_tokenize(raw)
```

Appendix E

```
lemmer = nltk.stem.WordNetLemmatizer()
def LemTokens(tokens):
    return [lemmer.lemmatize(token) for token in tokens]
remove_punct_dict = dict((ord(punct), None) for punct in string.punctuation)
def LemNormalize(text):
    return LemTokens(nltk.word_tokenize(text.lower().translate(remove_punct_dict)))
```

Appendix F

```
GREETING_INPUTS = ["اهلا و سهلا", "اهلا", "مرحبا"]
GREETING_RESPONSES=["اهلا و سهلا", "اهلا", "مرحبا"]

def greeting(sentence):
    for word in sentence.split():
        if word.lower() in GREETING_INPUTS:
            return random.choice(GREETING_RESPONSES)
```

Appendix G

```
def response(user_response):
    robo_response=''
    sent_tokens.append(user_response)
    TfidfVec = TfidfVectorizer(tokenizer=LemNormalize, stop_words='english')
    tfidf = TfidfVec.fit_transform(sent_tokens)
    vals = cosine_similarity(tfidf[-1], tfidf)
    idx=vals.argsort()[0][:-2]
    flat = vals.flatten()
    flat.sort()
    req_tfidf = flat[-2]
    if(req_tfidf==0):
        robo_response=robo_response+"لم أفهم ما هو المقصود"
        return robo_response
    else:
        robo_response = robo_response+sent_tokens[idx]
    return robo_response
```

Appendix H

```
flag=True
print("SeerahBot: انا مساعدك الذكي، استطيع الإجابة على أسئلتك حول السيرة النبوية")
while(flag==True):
    user_response = input("You: ")
    user_response=user_response.lower()
    if(user_response!='الى اللغاه'):
        if(user_response=='فكرنا' or user_response=='لك' or user_response=='فكرنا'):
            flag=False
            print("SeerahBot: غنوا..")
        else:
            if(greeting(user_response)!=None):
                print("SeerahBot: "+greeting(user_response))
            else:
                print("SeerahBot: ",end="")
                print(response(user_response))
                sent_tokens.remove(user_response)
    else:
        flag=False
        print("SeerahBot: !! الى اللغاه..")
```