REVISIT DIALOGFLOW IN AN ENGLISH TEACHING VIRTUAL ASSISTANT USE CASE

M.S. Tran[1], T.H. Tran[2], Q.D. Tran[3] and D.T. Nguyen[1]

[1]AI Lab, Topica Holding, Hanoi, Vietnam
[2]University of Technology and Education, Hochiminh City, Vietnam
[3]Hanoi University of Science and Technology, Hanoi, Vietnam

ABSTRACT

We deployed a conversation chatbot as a virtual assistant teaching English via Internet. The system was developed on the base of Moodle as Learning Management System and DialogFlow as Dialogue Management System. It is interesting that the crucial problem we had to face here is the lack of an efficient authoring tool in order to generate in mass the dialog scenarios fed into Moodle and DialogFlow. In our concrete case - teaching English for beginners - the Dialogflow platform seems to be a cumbersome tool. Especially with bad internet connection, sending messages back and forth to Dialogflow may degrade smooth conversation experience. We therefore built an authoring tool to fasten up the conversation rules generation. We also replace Dialogflow with a local browser-based dialogue management engine. The lessons taught with our systems – our English teaching virtual assistant – seem interesting to students and receive encouraging feedbacks.

KEYWORDS

Chatbot, Dialogflow. Artificial Intelligence, Natural Language Processing.

1. INTRODUCTION

Together with 4 industrial revolutions, we are the witness of 7 technology waves in education. Firstly, the emergence of the Internet indeed makes study available outside the wall of schools / universities. But only with the boom of smart phones, students can now really enjoy learning at anytime and anywhere if they just possess a small device – a smart mobile phone – having a computational power as a huge desktop computer of the 80s. The third wave of technology in education is the mixture between online and offline, namely Online to Offline (O2O) service model. The learners collaborate online then do (partial) practice of offline physical spaces to fulfill the study goal. The fourth wave focuses on the quality of the online teaching materials. With the simulation capability of computer, more natural and realistic teaching contents are now available thanks to 3-dimensional design, Augmented Reality, Virtual Reality technology. The fifth wave is the individual tutoring, where one can contact a privilege teacher via a video conference application. The 6th and 7th waves can be stated as the era of the Artificial Intelligence (AI). In the former, AI plays a role of supplemental tools to help teachers work more efficiently, while the latter addresses students as a central target for a personalized teaching plan.

Chatbot, especially conversational chatbot is a representative for the 6th wave of educational technology. There are a lot of popular conversational chatbots available today like Apple's Siri, Microsoft's Cortana, Google Assistant and Amazon's Alexa [7]. However chatbot dedicated to English teaching in a dialoguing manner following a strict curriculum is likely rare.

Natarajan Meghanathan et al. (Eds) : CoSIT, AIAPP, SIGL, CRIS, DMA, NLPML, CYBI - 2020
pp. 11-17, 2020. © CS & IT-CSCP 2020 DOI: 10.5121/csit.2020.100102

In this article, the architecture of a conversational chatbot will be introduced. The chatbot is used as an English Teaching Virtual Assistant (ETVA) in our Edtech Group to teach beginners who want to learn English. It is about an English curriculum having 10 units, each has four lessons of about 20 minutes of dialogue. Therefore, the deployment of a smooth and natural dialogue management system is obviously necessary, but the rapid process to setup such huge teaching materials / conversation situations is also a considerable factor.

The article is arranged as the following. In the next section, the general architecture of the ETVA will be analyzed. The third section is dedicated to an authoring tool to ease the building process of conversational scenarios related to English lessons. The fourth section points out some shortages of Dialogflow initially used as our dialogue management system. A simplified model of dialogue management, best fitted to our English teaching curriculum is designed and discussed in this section. The article is closed with conclusion and perspective in the fifth section.

2. ARCHITECTURE OF THE ETVA SYSTEM

Figure 1outlines the architecture of the ETVA. The core components of the system are Moodle [2] as the Learning Management System LMS and the Google's Dialogflow [1] as the dialogue management system. The former keeps track of all studying materials (learnt lessons, the grades, logs,..,) of students' study-path. The latter, a development platform for rule-based chatbot [7], determines the goodness of the virtual assistant. To achieve this goal, careful conversation scenarios must be fed into Dialogflow via its own Graphic User Interface (Figure 1). In fact, the Authoring Tool in Figure 1 is partly supported by Dialogflow itself.

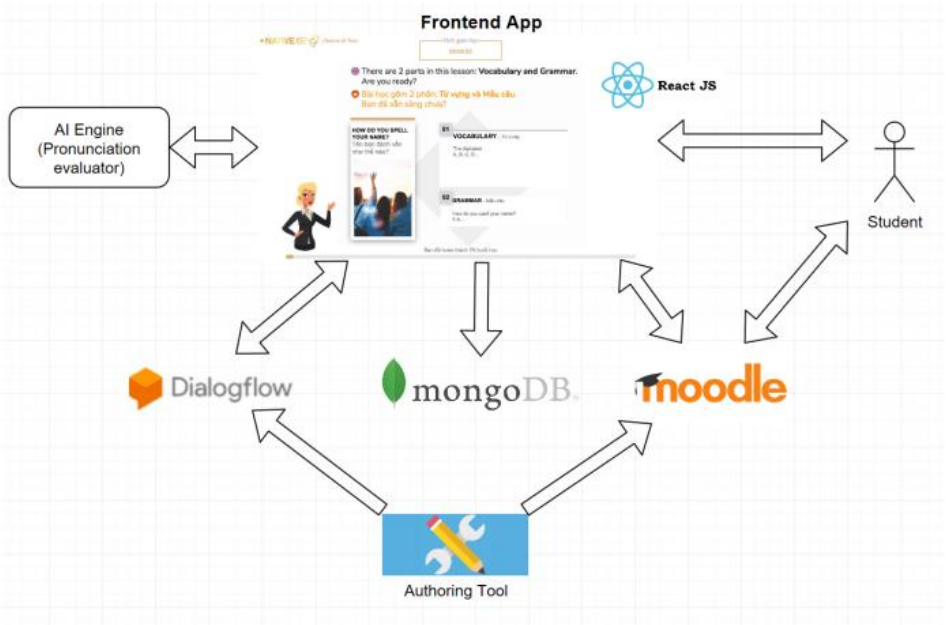


Figure 1: Architecture of the Virtual Lecturer Application.

Students and the virtual assistant talk / discuss to each other in accordance with presentations shown on a computer's screen just as a real teacher presents the presentation to his / her students. In the current version, 3 types of slides can be found in the presentation. A lecture-only slide contains uniquely texts to be read to student via the Artyom Speech2Text engine [11]. In case of the interactive slides, the virtual assistant askes student to interact with the contents on the slide according to English commands. Selecting an image, dragging and dropping graphic objects are currently supported as such interactions. The pronunciation slide is the third

type, where student is requested to pronounce a word / a sentence. In this case, the student's utterance will be evaluated with the *Pronunciation evaluator* Speechace [12] and at the same time be recorded to the *mongoDB* as in Figure 1. The slides are played back one after the other in accordance with the speaking rhythm of the virtual assistant. The behaviours related to 3 types of slides are implemented as an essential feature of *Frontend App* with React JS framework.

3. IMPORTING CONVERSATION LOGIC TO DIALOGFLOW

Dialogflow is a natural language understanding platform that makes it easy to design and integrate a conversational user interface into any application. Dialogflow can analyze multiple types of input from an application, including text or audio inputs. It then responds to the application in a couple of ways, either through text (can be used as parameters for further processing like rendering some special effects conditionally), image or with synthetic speech according to predefined rules given as a tree-graph of Intents [1]. Intents are activated when one of predefined textual / audio patterns of that intent is matched. Intent may have a pre- and post-context to reinforce the activation order of the intents in the graph.

According to our English teaching curriculum, one conversation / lesson consists of about 100 intents. Some intents (pronunciation and interactive slides) have post-contexts. Although Dialogflow has a user friendly GUI for composing these intents, the typing process is long and cumbersome. Furthermore, it is impossible to view the tree in fully collapsed mode in order to review all intents and their derived ones for the sake of verification.

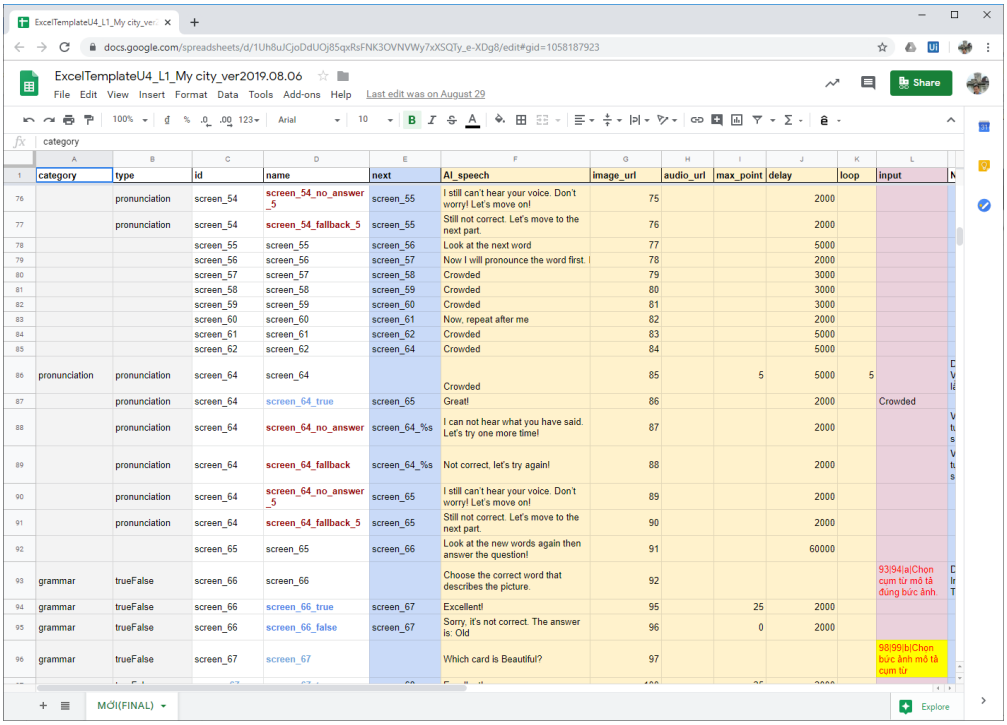
We found that two-level tree – cause and consequence - of intents 'graph can be straightforwardly edited and checked in a table-wise document like in an excel or google sheet format. Hereafter the term "slide" and "intent" are exchangeable. It is because each intent in Dialogflow is uniquely assigned to a slide on our lesson.

Figure 2 illustrates a table-fashion normalized data structure of the intents in one lesson. The column L contains the textual pattern for the given intent (given row). Only intent corresponding to pronunciation slide has one or more value (multi correct patterns) in this column. Columns from F to K (yellow area) consist of several variables that the intent will send back to the application *Frontend App*. For instance, the *Frontend App* (Figure 1) will use the text found in column F to drive the Text-to-Speech to generate synthetic voice for the associated slide identified as the value in column C. In parallel, the *Frontend App* can show the image having the index in column G. Then it will pause for a certain milliseconds as specified in the column J before calling the next intent in the column E. For the intent of type *trueFalse* or *pronunciation* (value in column B), the value in column D will be the conditional intent according to the input from student.

Let's look at the intent of type *trueFalse* in the row 93. The value in the Column L – "93|94|a| Chọn cụm từ mô tả đúng bức ảnh" – in this case implies two images having index 93 and 94 respectively. Receiving this value from *Dialogflow*, the *Frontend App* will show up these images with the title "Chọn cụm từ mô tả đúng bức ảnh" – the string after the last separator "|". If the student selects the first one (referred to as the index "a" in "93|94|a|Chọn cụm từ mô tả đúng bức ảnh") the *Frontend App* will call the intent *screen_66_true* - a post-context intent of the intent *screen_66*. Otherwise – the student selects the wrong image i.e. image 94 - the *Frontend App* call the intent *screen_66_false*.

Using the above described excel format, we can easily edit, maintain and verify the curriculum throughout the lessons. To avoid manually transferring data from excel to Dialogflow, we

implement an importing service, automatically export the information from excel to a compressed format, which can be directly imported into Dialogflow. As a result, editing a lesson having about 100 intents becomes an instantaneous service in a range of several minutes. Before, directly using Dialogflow GUI to feed the same content takes us about 4 to 5 working hours with higher typing error probability.



	category	type	id	name	next	AI_speech	image_url	audio_url	max_point	delay	loop	input
76		pronunciation	screen_54	screen_54_no_answer_5	screen_55	I still can't hear your voice. Don't worry! Let's move on!	75			2000		
77		pronunciation	screen_54	screen_54_fallback_5	screen_55	Still not correct. Let's move to the next part.	76			2000		
78			screen_55	screen_55	screen_56	Look at the next word	77			5000		
79			screen_56	screen_56	screen_57	Now I will pronounce the word first.	78			2000		
80			screen_57	screen_57	screen_58	Crowded	79			3000		
81			screen_58	screen_58	screen_59	Crowded	80			3000		
82			screen_59	screen_59	screen_60	Crowded	81			3000		
83			screen_60	screen_60	screen_61	Now, repeat after me	82			2000		
84			screen_61	screen_61	screen_62	Crowded	83			5000		
85			screen_62	screen_62	screen_64	Crowded	84			5000		
86	pronunciation	pronunciation	screen_64	screen_64		Crowded	85		5	5000	5	
87		pronunciation	screen_64	screen_64_true	screen_65	Great!	86			2000		Crowded
88		pronunciation	screen_64	screen_64_no_answer	screen_64_%s	I can not hear what you have said. Let's try one more time!	87			2000		
89		pronunciation	screen_64	screen_64_fallback	screen_64_%s	Not correct, let's try again!	88			2000		
90		pronunciation	screen_64	screen_64_no_answer_5	screen_65	I still can't hear your voice. Don't worry! Let's move on!	89			2000		
91		pronunciation	screen_64	screen_64_fallback_5	screen_65	Still not correct. Let's move to the next part.	90			2000		
92			screen_65	screen_65	screen_66	Look at the new words again then answer the question!	91			60000		
93	grammar	trueFalse	screen_66	screen_66		Choose the correct word that describes the picture.	92					93|94|a|Chọn cụm từ mô tả đúng bức ảnh.
94	grammar	trueFalse	screen_66	screen_66_true	screen_67	Excellent!	95		25	2000		
95	grammar	trueFalse	screen_66	screen_66_false	screen_67	Sorry, it's not correct. The answer is: Old	96		0	2000		
96	grammar	trueFalse	screen_67	screen_67		Which card is Beautiful?	97					98|99|b|Chọn bức ảnh mô tả cụm từ

Figure 2: Excel format for dialogue scenario.

4. LOCAL DIALOG MANAGEMENT SYSTEM A MORE RESPONSIBLE CONVERSATION

In our specific case: teaching English for beginners, the conversation scenario is quite simple. A lesson of 100 slides in average only has about 5 slides for pronunciation and interaction. These two types essentially require dialogue management. In addition, the conversation scheme does not go further than the 2nd level of the tree graph. For the majority of the slides, they just need to be played back successively without conditional branch. Other factor is that the request-answer circulated between *Frontend App* and *Dialogflow* may reduce the experience of natural chattering with teacher (a virtual one) due to fluctuation of Internet bandwidth.

It is the main reason motivating us to implement our own dialog management system in the place of *Dialogflow*. The new dialog core – hereafter referred to as Local Dialog Manager LDM - is embedded locally in the *Frontend App*.
Figure 3 summarizes some principles of the LDM.
The LDM process is in fact a loop through all the slides (injective association with intents in Dialogflow). Each slide is classified into 3 types: *Lecture*, *Pronunciation* or *Interactive*, which are shown in the left, middle and right branches respectively (
Figure 3).

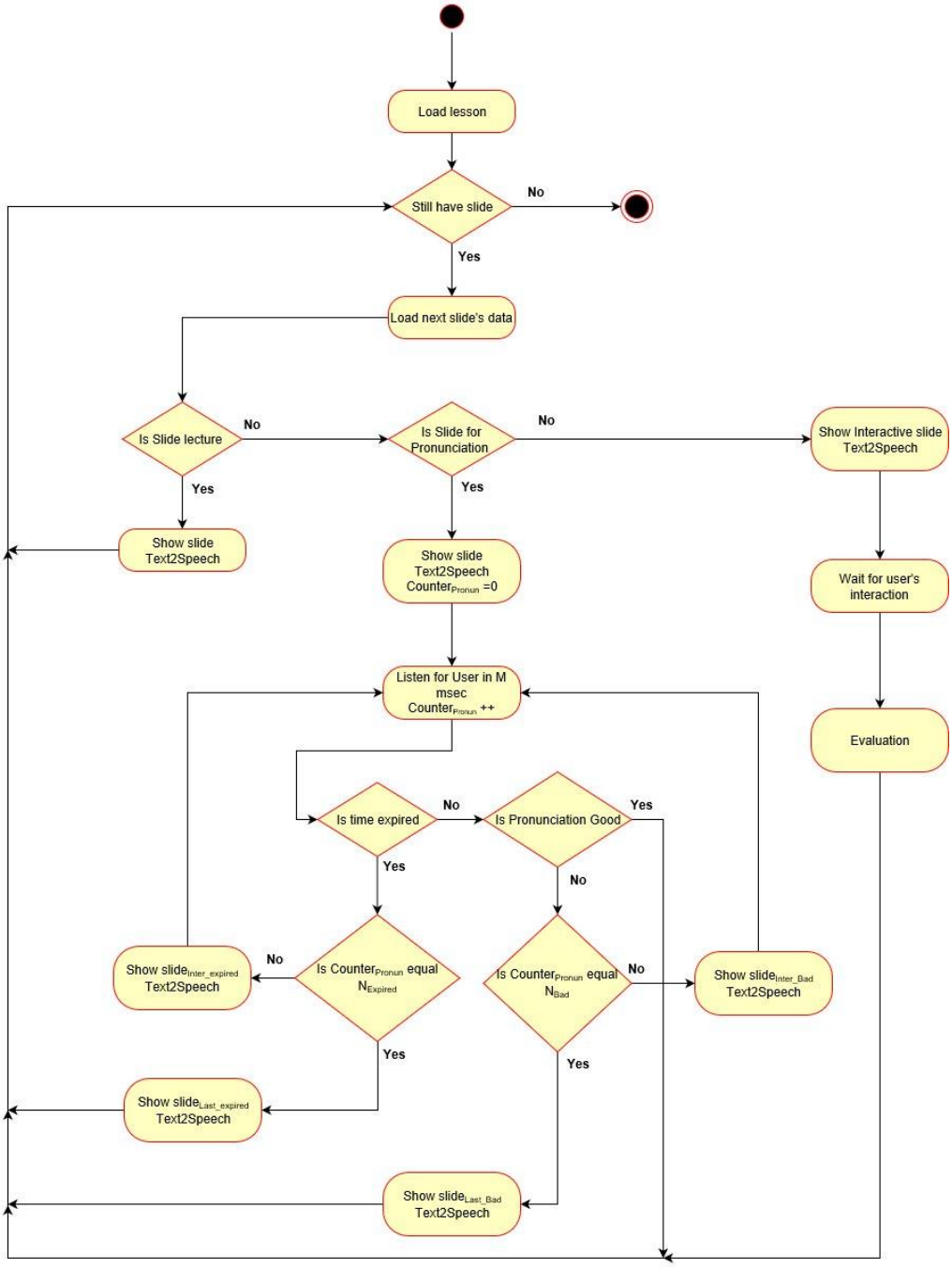


Figure 3: Local Dialog Management flow of the ETVA

4.1. Lecture slides

If the slide is of type lecture, the LDM drives the *Frontend App* to show the images, to synthesize the texts, to delay for a certain amount of time and to move to the sequential slide exactly as described for the slide in the excel format. The life-cycle of this type is shown on the left branch in
Figure 3

4.2. Pronunciation slides

In case of *Pronunciation*, a student is expected to pronounce a certain text pattern. Consequently there are three predefined sub-branches designed in LDM:

I. The student say nothing after a given waiting period,

II. The student say incorrectly, and

III. The student say correctly

The block *Listen for User in M msec* (
Figure 3) is the state of LDM while waiting for student's utterance. In the branch *Yes* of the block *Is time expired* (case I), a prompt slide (the block *Show slideInter_expired*) is shown to permit the student retry up to *Nexpired* times. In the last try, the block *Show slideLast_expired* is activated, and the LDM redirects to the next slide.

In the branch *No* of the block *Is time expired* continuing to branch *No* of the block *Is Pronunciation Good* - bad pronunciation as in case II - the block *Show slideInter_Bad* will let the student retry the practice. If the student continues pronouncing incorrectly up to *NBad* times, the state *Show slideLast_Bad* is triggered, and the LDM moves to the next slide.
In case III (the branch *No* of the block *Is time expired)* then the branch *Yes* of the block *Is Pronunciation Good* the state *Show slideCongratulation* is activated. The next slide is handled afterward.

4.3. Interactive slides

The LDM falls into this state along the path *No* of the block *Is Slide lecture* and *No* of the block *Is Slide for Pronunciation*. LDM will order the *Frontend App* to display the selectable images and wait for the student's input. Upon the student's final choice, the *Evaluation* state takes care of the promoting message and redirects to the next slide.

In Figure 3, most of the states / blocks have the notation *Speech2Text*, which implies the voice synthesizing operation occurring in all slides according to our pedagogy plan.

Although we lost the flexibility of a generic purpose dialogue management as well as the feature of natural language understanding integrated in *Dialogflow*, making use of LDM provides us a compact but best-fit to our teaching target. The quality of experience is also improved with the responsiveness of the system.

5. CONCLUSION AND PERSPECTIVES

Going together with latest waves in educational technology, notably applying AI as assistant tools in education, we design our own chatbot ETVA to facilitate the English teaching for beginner. To reduce the gap between real teacher and virtual one, we carefully analyze each component of the system. Our class' frontend interface, like many other online teaching applications, is designed to be rich in multimedia content. The conversational capability further makes the interface friendlier and especially more human. To ensure the quality of experience – the possibly most natural conversation - while communication with our ETVA, we customized the *Dialogflow* logic to best fit to our pedagogical plan. The responsiveness of the system is also taken into consideration. The last but not least is the process to build the teaching content in an

efficient manner. We setup a procedure end-to-end from outlining, editing, verifying and committing teaching material / scenarios to ETVA so that the teaching content can be created in the rapidest and least erroneous manner.
In the future, we will extend English teaching Assistant to higher level of student, requiring more intelligent conversation. We also try to improve the responsiveness, and hence the naturalness of dialoguing with ETVA via more precisely recognizing the end of student's utterances.

Our ETVA is being piloted with students of our education center, the impression and feedback upon the real application of learning with virtual teacher is encouraging.

ACKNOWLEDGEMENTS

We would like to thank Topic a Holding Group to define the scope as well as to fully fund the work. We express our gratitude to the AI Lab staff, who dedicated their skills and energies to make the virtual teacher concept to be deployed in the real teaching environment at Topica.

REFERENCES

[1] Dialogflow documentation. Official website https://cloud.google.com/dialogflow/docs/

[2] Moodle Docs 3.7 2019. Official website: https://moodle.org/

[3] Shawar, Bayan Abu and Eric Atwell 2007. Chatbots: are they really useful?. In: Ldv forum. Vol. 22. 1, pp. 29–49.

[4] Ranoliya, Bhavika R, Nidhi Raghuwanshi, and Sanjay Singh 2017. Chatbot for University Related FAQs. International Conference on Advances in Computing, Communications and Informatics (ICACCI). Udupi, pp. 1525–1530.

[5] Monica Gill 2019. 5 ways Artificial Intelligence and Chatbots are changing education. https://towardsdatascience.com/5-ways-artificial-intelligence-and-chatbots-are-changing-education-9e7d9425421d

[6] Sofie Roos 2018 Chatbots in Education a passing trend or a valuable pedagogical tool, https://pdfs.semanticscholar.org/533e/bc0255c36749e1f46b8d3662464d6ee5d4f0.pdf

[7] Kunal Bhashkar, 2019. Build your own chat bot https://medium.com/@BhashkarKunal/conversational-ai-chatbot-using-deep-learning-how-bi-directional-lstm-machine-reading-38dc5cf5a5a3

[8] Branislav Srdanovic, 2017. Chatbots in Education: Applications of Chatbot Technologies. https://elearningindustry.com/chatbots-in-education-applications-chatbot-technologies

[9] Sundar Krishnan, 2019. How to use Google Speech to Text API to transcribe long audio files. https://towardsdatascience.com/how-to-use-google-speech-to-text-api-to-transcribe-long-audio-files-1c886f4eb3e9

[10] Maruti techlabs. 14 most powerful platforms to build a Chatbot. https://marutitech.com/14-powerful-chatbot-platforms/

[11] Speech2Text SDK. Official website https://sdkcarlos.github.io/sites/artyom.html

[12] Speechace Developer API Documentation, https://docs.speechace.com/?version=latest