



Journal Homepage: -www.journalijar.com

INTERNATIONAL JOURNAL OF ADVANCED RESEARCH (IJAR)

Article DOI:10.21474/IJAR01/15886
DOI URL: <http://dx.doi.org/10.21474/IJAR01/15886>



RESEARCH ARTICLE

A MODEL OF A VIRTUAL TUTOR CAPABLE OF SEEKING THE ANSWER TO THE LEARNER'S CONCERN IN THE TEACHER'S COURSE MATERIAL

Mamadou Bakouan, Yao Konan and Kone Tiémoman

Department of Computer science and Digital Sciences; Research Unity of Digital Expertise; Virtual University of Côte d'Ivoire.

Manuscript Info

Manuscript History

Received: 15 October 2022

Final Accepted: 18 November 2022

Published: December 2022

Keywords:-

Virtual Tutor, E-Learning, Chatbot,
Natural Language Processing

Abstract

In the operation of online and distance learning institutions, tutors are responsible for coaching, mentoring, and guiding learners during their courses. However, when a learner needs a tutor, the tutor is sometimes overwhelmed and unable to satisfy the learner in time. Moreover, the mode and time of work differ from student to student, which makes monitoring very difficult. Thus, to overcome the unavailability of a tutor or teacher, we proposed a conversational agent that can address the concerns of students in real time based on the teacher's course. The objective was to design a chatbot that can interact with the learner, understand his/her concerns and provide accurate answers at any time, autonomously browsing the course of the concerned teacher. We used a natural language processing (NLP) method to enable our chatbot to understand human language and respond to questions asked. The results showed that our chatbot manages to provide an adequate response depending on the context.

Copy Right, IJAR, 2022, All rights reserved.

Introduction:-

During the Corona crisis and the subsequent lockdown, we learned to do many tasks at home such as shopping for food, clothing, and more. This period also encouraged distance learning called online courses. Indeed, distance learning means that instead of being in a classroom and having a teacher giving us lessons at a specific time, we take our lessons without constraints of time and place, and we can view the course as much as we want with the help of human-computer interaction technologies (Caraban, 2019). Indeed, today, communicating with your machine in natural language is more than necessary. Recent advances in technology and artificial intelligence have led to the emergence of interactive systems in almost every field. For (Bei, 2022), a chatbot is a virtual agent that can interact with users using artificial intelligence (Dick, 2019). This agent can also perform tasks such as responding quickly to a user, informing them, helping them purchase products, and providing customers with better services (Xu, 2017). In recent years, several chatbots have emerged, such as Apple's SIRI conversational agent, and Amazon's Alexia. These chatbots make our lives easier and are on their way to winning all domains. In the field of education, especially distance learning, chatbots could have huge potential. Indeed, in this type of education, each learner must have a specialized tutor. However, this type of service is often unavailable or when it exists, exchanges between tutors and students are sometimes difficult. Tutors often answer the same questions and students sometimes complain about the time it takes the tutor to respond to their concerns. In fact, during study hours the student may have questions about a part of the course, it would be ideal for him/her to find a clear answer to these questions to better understand the course and move forward quickly. But often at that very moment, the teacher has either taken a

Corresponding Author:- YAO Konan

Address:-Department of Computer science and Digital Sciences; Research Unity of Digital Expertise; Virtual University of Côte d'Ivoire.

break, is busy, or is even unwell. Also, a human tutor may have more than one student, which makes communication slow and tedious. So, what is the most practical and economical alternative that can be found to address the learner's concerns quickly and accurately?

Generally, chatbots are used as virtual assistants to solve customer service problems at the right time (Liu, 2020; Nuruzzaman, 2018). However, they can be used to address the concerns of e-learners (Mohd, 2022) in specific situations. Indeed, chatbots can identify a user's intent, extract relevant data from his or her request and provide the most appropriate response. In this paper, we propose a chatbot model that can understand learners' concerns, search for the answer in the teacher's course and provide the most appropriate response. This approach is based on the Natural Language Processing (NLP) method. The main contributions of this work are: to propose a tool capable of understanding the student's concerns and supporting him 24 hours a day; Design a virtual agent capable of reading a pdf file.

The rest of this work is constituted as follows: in section 2 we present the materials and methods used, and the third section is devoted to the results and discussions. Finally, in section 4 the conclusion.

Materials and Methods:-

Chatbots

Most of the time, called "bots", these computer programs can understand and process natural language, to communicate with a user through text or voice messages. The bot or conversational agent (Adamopoulou, 2020) is a virtual agent capable of giving an adequate answer to a question posed by a human user. It is a system that acts like a human being and helps a user to accomplish a task. Due to their ability to mimic human conversations and automate services, chatbots are increasingly being used in several areas. Sometimes they are used in the field of training (Martin, 2022), health (Chang, 2021), commerce as an advisor (Xu, 2017), or even as an academic supervisor (Kuhail, 2022). In higher education, chatbots are used to understand student behaviors and to personalize learning to avoid learner dropouts (Mohd, 2022).

Corpus

The first step during the design is the recovery of the corpus. The corpus is a set of documents or an aggregate of text grouped in a specific theme. This is the set of data that we will have to process. We must read the files in pdf format for this work because the teachers' lessons are generally in this format to facilitate asynchronous access. Thanks to the "pdfx" library, we retrieve the file via a link using the "get_pdf_file" function. We copy the content of the page and put it in a file named "pdf_file".

```
!pip install pdfx8
```

Natural Language Processing

Natural Language Processing (NLP) is the branch of artificial intelligence that aims to give machines the ability to understand human language as spoken and/or written, generate it, and translate it. The NLP aims to classify text and recognize text (Bansal, 2018), make summaries (Caraban, 2019), and automatic translation (Berroubi, 2022), etc. Generally, NLP models are subdivided into two main categories, in particular, machine learning models with natural language understanding (NLU) (Bedjaoui, 2022) which focus on understanding speech in context, and machine learning models oriented natural language generation (NLG) (Berroubi, 2022) which focus on how to generate text the way humans do.

Pre-processing text with NLTK

The Natural Language Tool Kit (NLTK) is a python language-interfaced library integrating a wide range of language analysis algorithms and methods. It thus offers different methods of tokenization, annotation in parts of speech, chunking, detection of named entities, syntactic analysis, etc. Indeed, the main problem with textual data is that they are all in text format (character strings). Yet, machine learning algorithms need some kind of numerical feature vector to accomplish their task. Thus, basic text preprocessing consists of preparing and cleaning the textual data so that the machine can analyze it. It is also, in our case, the transformation of PDF files into TXT, followed by the conversion of the entire text into a single case (upper or lower case), here in lowercase, so that the algorithm does not process the same words in cases different from different. We install the spacy library: « pip install spacy » for analysis of the text. First, we install the following extensions for processing in French.

```
!python -m spacy download fr_core_news_sm
```

```
!python -m spacy download en_core_news_lg
Installing NLTK packages
```

```
!pip install nltk
```

Tokenization:

The text is broken down into small units to work with at this level. Text tokenization or vectorization refers to the process of converting text strings into a list of tokens. The phrase tokenizer can be used to find the list of phrases and the word tokenizer can be used to find the list of words. With NLTK, text can be split by words with the `word_tokenize()` or by sentences with the `sent_tokenize()` function.

```
import nltk
nltk.download('punkt')
```

We used the Spacy tokenizer. The Spacy tokenizer is a fast tokenizer that allows the user to specify special tokens that will not be tokenized or that will be tokenized in a specific way.

Stop Word :

Certain words such as articles, pronouns (Je, il, ...), etc. are common when formulating a search query but have little importance or meaning in the search. Thus, it is a question of removing these words from the text that the unique words offer more information at the request of the learner remain.

```
import nltk
nltk.download('stopwords')
```

Lemmatization:

The lemmatization process consists of the creation of a lexical database. Concretely, it is a question of reducing words to their root forms, for example, putting verbs in the infinitive, and removing agreements (messages → message). In the script below, we define a function that handles the conversion of text to lowercase, down to the removal of numbers, some special characters, and spaces. This function gives us a list of words.

```
def clean(text):
    return [ unicode(token.lemma_.lower()) for token in docx
             if (not token.is_punct)
                and (not token.is_space)
                and (not token.like_url)
                and (not token.is_stop)
                and len(token) > 1
                and (not token.like_num)]
```

TF-IDF is an abbreviation for Term Frequency (TF) - Inverse Document Frequency (IDF). The TF is the frequency with which a given term is used in a document and IDF refers to the importance of a given term in a set of documents. This approach makes it possible to determine in what proportions certain words of a text document or a body of a document can be evaluated compared to a given text. TF-IDF vectorization involves calculating the tf-idf score for every word in your corpus relative to that document and then putting that information into a vector.

$$tf = \frac{\text{Number of occurrences of a word}}{\text{Total words in the document}}$$

$$idf(\text{word}) = \log\left(\frac{\text{Total number of documents}}{\text{Number of documents containing the word}}\right)$$

By multiplying TF and IDF together we can get our final TF-IDF value.

$$tf.idf(t, d, D) = tf(t, d).idf(d, D)$$

where t is the term (word), d the documents in the corpus (D). TF-IDF is a transformation applied to texts to obtain two vectors. We can then obtain the similarity cosine between the two documents (d_1, d_2).

$$\text{similarity}(d_1, d_2) = \frac{d_1 \cdot d_2}{\|d_1\| \|d_2\|}$$

Results And Discussions:-

In this method, a user (learner) asks a question via the user interface, his request is processed by the chatbot in the first place by the language comprehension module which makes it possible to extract data from words to understand their meaning. Then the answer to the request is sought in the teacher's course which is in a PDF file. The response is produced via the natural language generation module.

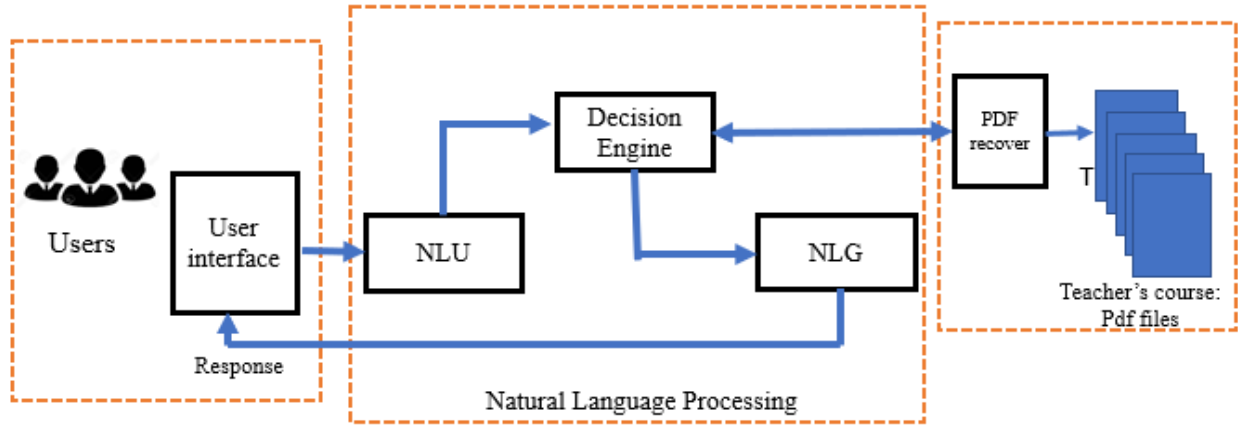


Figure1:- Process architecture.

Exchangetests with our Virtual tutor

the learner wants to understand the word “programme”
 MynameisTutor_bot and I'm a chatbot, I amhere to help you.

You: programme

Totor_bot:L'ensemble des instructions s'appelle un programme. Le programme appellant donne au sous-programme les valeurs des contenus des variables (ou des expressions) utilisées comme paramètres. Cependant, les variables déclaa- rées dans le programme principal sont distinctes de celles utilisées dans un sous-programme.

You :

the learner wants to understand the word “ scanner”.

You: scanner

Totor_bot: Scanner à Plat Tous les dispositifs FAX utilisent un scanner
 Un scanner (dans le milieu informatique) est un appareil qui permet de digitaliser un document.

You: ok

Tutor_bot: Can I tell you more about it

You: bye

Tutor_bot: Ok, Thanks, Bye...

Stay Home,and take care of yourself!

To ensure the proper functioning of our program, it is important to carry out tests. we have provided as input a learner concern. As output, the chatbot provides information about learner concerns that are in the teacher's course.We notice that our chatbot manages to provide appropriate answers taking into account the context as well as the similarity of the words. the measure of similarity with the function, the similarityof the spacy library gives the following similarity scores (table 1). This result opens a new perspective for smarter chatbot design by integrating context analysis modules.

Table 1:- Similarity score between Learner concern and bot response.

Learner concern	Bot response	Spacy similarity score
Programme	L'ensemble des instructions s'appelle un programme. Le programme appellant donne au sous-programme les valeurs des contenus	0.3723546891488248

	des variables (ou des expressions) utilisées comme paramètres. Cependant, les variables déclarées dans le programme principal sont distinctes de celles utilisées dans un sous-programme.	
Scanner	Scanner à Plat Tous les dispositifs FAX utilisent un scanner Un scanner (dans le milieu informatique) est un appareil qui permet de digitaliser un document.	0.33070704972102394

Conclusion:-

From the above study, we proposed a virtual tutor able to seek the response to the learner questions autonomously in the teacher's course. It contributes to improving the follow-up of learners in higher education who take online and distance learning courses. Through this virtual agent, learners can freely ask their concerns at any time and get the answer. With this approach, the learner's response is provided taking into account the context in which the words are used in the query.

In the coming months, we plan to deploy our chatbot in the virtual university of Côte d'Ivoire follow-up system for improving the existing chatbot. We also want to provide a new version by integrating a context-processing module into the design process.

References:-

1. Adamopoulou, E. & Moussiades, L (2020). An overview of chatbot technology. in IFIP International Conference on Artificial Intelligence Applications and Innovations 373–383 (Springer, 2020).
2. Bansal, H. & Khan, R (2018). A review paper on human-computer interaction. International Journals of Advanced Research in Computer Science and Software Engineering **8**, 53–56.
3. Bedjaoui, N. (2022). Corpus linguistics, TAL, and machine translation.
4. Berroubi, a. & ben latrache, s. (2022). Towards a system for the automatic summarization of Arabic texts. (University of m'sila,).
5. Cahn, J (2017). CHATBOT: Architecture, design, & development. University of Pennsylvania School of Engineering and Applied Science Department of Computer and Information Science.
6. Caraban, A., Karapanos, E., Gonçalves, D. & Campos, P (2019). 23 ways to nudge: A review of technology-mediated nudging in human-computer interaction. in Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems 1–15.
7. Chang, C.-Y., Hwang, G.-J. & Gau, M.-L (2022). Promoting students' learning achievement and self-efficacy: A mobile chatbot approach for nursing training. British Journal of Educational Technology **53**, 171–188.
8. Chen, Y., Jensen, S., Albert, L. J., Gupta, S. & Lee, T (2022). Artificial Intelligence (AI) Student Assistants in the Classroom: Designing Chatbots to Support Student Success. Information Systems Frontiers 1–22.
9. Dahiya, M. A tool of conversation : Chatbot (2017). International Journal of Computer Sciences and Engineering **5**, 158–161.
10. Dick, S. (2019). Artificial intelligence. Publisher: PubPub.
11. Kuhail, M. A., Alturki, N., Alramlawi, S. & Alhejori, K (2022). Interacting with educational chatbots: A systematic review. Education and Information Technologies 1–46.
12. Liu, C., Jiang, J., Xiong, C., Yang, Y. & Ye, J (2020). Towards building an intelligent chatbot for customer service: Learning to respond at the appropriate time. in Proceedings of the 26th ACM SIGKDD international conference on Knowledge Discovery & Data Mining 3377–3385.
13. Luo, B., Lau, R. Y., Li, C. & Si, Y.-W (2022). A critical review of state-of-the-art chatbot designs and applications. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery **12**, e1434.
14. Martin, C. & Li, Y. A Scalable Self-Learning Chatbot for Classroom (2022). In 2022 IEEE 12th Annual Computing and Communication Workshop and Conference (CCWC) 0915–0921.

15. Mohd Rahim, N. I., A. Iahad, N., Yusof, A. F. & A. Al-Sharafi, M(2022). AI-Based chatbots adoption model for higher-education institutions: A hybrid PLS-SEM-Neural network modelling approach. *Sustainability***14**, 12726.
16. Nuruzzaman, M. & Hussain, O. K (2018). A survey on chatbot implementation in customer service industry through deep neural networks. in 2018 IEEE 15th International Conference on e-Business Engineering (ICEBE) 54–61.
17. Xu, A., Liu, Z., Guo, Y., Sinha, V. & Akkiraju, R(2017). A new chatbot for customer service on social media. in Proceedings of the 2017 CHI conference on human factors in computing systems 3506–3510.