International Journal of Multidisciplinary Research and Growth Evaluation

International Journal of Multidisciplinary Research and Growth Evaluation
ISSN: 2582-7138
Received: 28-05-2021; Accepted: 16-06-2021
www.allmultidisciplinaryjournal.com
Volume 2; Issue 4; July-August 2021; Page No. 146-151

Exploring use of Google speech recognition API, assisting teaching-learning of French language with special reference to UDUS

Abubakar Hassan [1], Zaki, Muhammad Zayyanu [2]
[1] Department of Computer Science, Faculty of Science, Usmanu Danfodiyo University, Sokoto, Nigeria
[2] Department of Modern European Languages and Linguistics, Faculty of Arts and Islamic Studies, Usmanu Danfodiyo University, Sokoto, Nigeria

Corresponding Author: **Abubakar Hassan**

Abstract

This paper titled *“Exploring the use of Google Speech Recognition API, Assisting Teaching-Learning of French Language with Special Reference to UDUS”* relates to current technologies such as Google Speech Recognition which is an Internet web-based software with the help of API which identifies, processes and manages articulated sound or words online. We are going to investigate its usage and application systematically in Google. This paper explores the use of the web-based Speech Recognition software designed that assists an easy teaching-learning process of the French language. Special emphasis is laid on the importance and use of Google APIs. Thus, the findings show that although the system is facing some challenges, it is very important in this contemporary period for ICT to dominate the teaching-learning process due to its vast application and capacity. In the online flow of application architecture, API serves as a messenger between clients and systems.

Keywords: Google, French language, Web-based software, API, Speech recognition, teaching-learning

Introduction

Interactions are achieved between human beings through a medium of communication, language. The interactions make humans unique creations in the world, especially with technological advancement - virtually. This transforms through the use of an interface. This interface is a messenger between users and machines. In general terms, technologies are collections of tools. Some of them affect our communications, and thus translation. It can be the units both substantial and unsubstantial manipulated by the application of mental and physical strength to attain some desired results, (Zaki, *et al*., 6). In this contexts, technology refers to tools and machines that may be used to solve real-world problems. In addition, such “tools and machines need not be material; virtual technology such as sophisticated software” [1]. The tool can be a simple one that aids in the teaching-learning process.

In this context, the French language is not left behind when it comes to globalization, as French and globalization are at the forefront. French is also the language of the Internet with more than 180 million users, it is ranked 6th by the number of visited sites per month, the 5th language of Wikipedia and the 3rd language on the Amazon after English and German, French is the third most used language on the Internet.

French language through communicating with French speakers – Francophones and Francophiles across the globe offer an alternative view of the world, and from its news body as the leading French-language international media like Radio France International (RFI) can be achieved online. The French language is made structurally an easy language to learn, speak and enjoyable for learners at all levels. It will not take time for the learner to reach a certain level of fluency.

Therefore, technology is more than human efforts as it is a tool that is needed to achieve a result. It is the state of humanity’s knowledge of how to combine resources to produce desired results through solving problems. This can be achieved through technical methods or skills, processes, techniques and tools, (Zaki., *et al*, 2). Hence, Information and Communication Technology (henceforth, ICT) encourages the learners of French to apply the skills obtained through ICT to enhance their performance in the language in the modern global age. Virtual or e-learning-teaching are made available that learners of French can only explore them when they possess adequate ICT skills, (Zaki, 3). Having ICT skills give the user a kind of self-confidence in operating

[1] Industry, Technology and Global Marketplace: International patenting Trends in two Technology Areas “Science and Engineering” indicators, 2002.

ICT gadgets and online operations.

Relevance of Google Speech Recognition in Teaching - Learning French

Globally, according to the French Ministry of Foreign Affairs and Europe (2016), more than 2% of the world's population speak French on the five continents. It comprises the Francophonie, the international organization of French-speaking countries, 68 States and Governments. As surveyed, French is the second most widely learned foreign language after English, and the ninth most widely spoken language in the world. French is also the only language, alongside English, that is taught in every country in the world. France operates the biggest international network of cultural institutes and centers, which run French-language courses for more than 750,000 learners. It is added that the "introduction of speech technology in the process of language teaching has been proven to be effective", (Hincks, 10).

Besides, an adaptation of speech technology enables outside-the-classroom language speaking practice by beginning language learners. Also, dialogue-based software which is another speech technology that uses fixed-response and Automated Speech Recognition enables learners to have a conversation with a computer by simulation. Learner's fluency and confidence get enhanced through practicing with such programs. Moreover, some of this software has an individual feedback mechanism that enables a learner to get feedback on pronunciation which is apparently lacked in the language classroom, (Rebecca, 6)

So, integrating speech recognition technology with Computer-Aided Language Learners (henceforth, CALL) activities recently became an interesting phenomenon to educators and CALL developers. Daniels *et al*, [4] reported that Speech recognition software was utilized before in Dyned's language learning software, in Subarashii, an interactive dialog system for learning Japanese and in echos, a voice interactive French language training system. They further report that today's robust mobile networks give speech recognition engines the ability to processes speech on powerful cloud servers. The smartphone device stands in as a microphone that sends the audio out over the Internet to a server which runs the CPU-intensive processing of the speech and sends back the transcribed text to the mobile device, (Daniels, *et al*., 6).

Furthermore, smartphones are now playing a vital role in solving the problem digital inequality, virtually in every part of the globe including remote areas smartphone has dominated not only communication but also both formal and informal learning process. Liao *et al*., (6) pointed out that the term "digital inequality" is often cited as "digital divide, "is gradually becoming history as digital devices are being increasingly integrated into the everyday lives of people around the globe both in rural and urban cities. Even though we cannot argue that there is still a wide margin in internet access and digital literacy between urban and rural communities. This might also be attributed to the non-availability of wired connections in the rural areas as it is in the urban areas. But since smartphones are being increasingly used for online access and the cost of smartphones is less compared to computer systems, especially Android devices. It has increased the capabilities of mid-range devices and brings digital resources and online language learning to long-underserved populations. This aids in bringing the desired changes to language teaching-learning and overall educational literacy occurring outside formal settings, (Robert, 12).

Thus, an adaption of this technology in the teaching-learning process becomes rapid due to the Covid-19 pandemic. Adeyinka-Ojo *et al*, [2] reported that "Covid-19 pandemic and technology advancement makes many higher educational institutions explore online teaching and learning platform" and it increases learners' enrolment. Usmanu Danfodiyo University was not left behind in this swift change, as an academic institution of learning that recently launched a Learning Management System to this effect. Therefore, exploring technologies such as Google Speech Recognition in teaching-learning of French becomes necessary to meet up to the challenge of e-learning as well as ensuring quality in the teaching-learning process.

The Concept of Google Speech Recognition

Speech Recognition

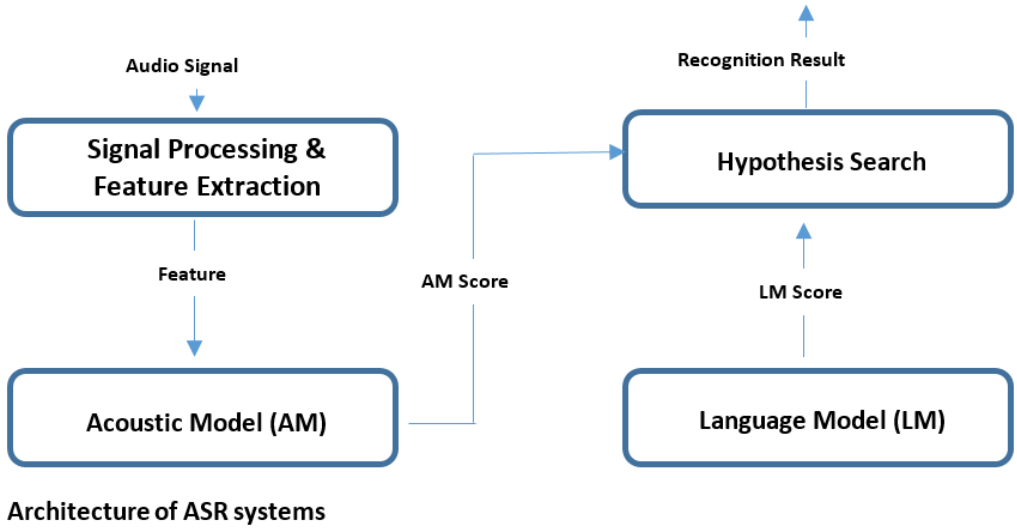
Speech Recognition is also referred to as "Speech-to-Text", which accurately converts speech into text or audio using API powered by Google's AI technologies. This facilitates the teaching-learning process as well as how technology can develop to advance learning languages like French. The technology recognizes speech allowing a voice to serve as the "main interface between the human and computer" [2]. The interface transcribes your content in real-time or from stored files and delivers a better experience in products through speech command. It has some features that support learning such as speech adaptation, streaming speech recognition, use of web vocabularies, multimodal recognition, noise robustness, content filtering, automatic punctuation (beta) and speaker diarization (beta).

Speech recognition has been described by Nenny *et al*., [6] as "the process of voice identification based on the spoken word by performing a conversion of a signal, this is captured by the audio device such as microphones". Sound is created by the act of vibration by something, this, in turn, sends a wave of energy into our ears, while speech is created by the sound created through spoken words. Speech recognition is an act of identifying sound for effective utilization. Speech Recognition can also be described as a system of capturing and recognizing word commands uttered mostly through human speech which is later translated into a form that can be acted upon by a computer. Speech recognition is divided into four approaches, this consist of acoustic-phonetic oncoming, artificial intelligence oncoming, a pattern recognition approach and a Neural Network approach. The latter has been in use by Automatic Speech Recognition (henceforth, ASR).

According to Nenny *et al*., [8] ASR system architecture has been used in different applications. It consists of four (4) components: signal processing and feature extraction, Acoustic Model (henceforth, AM), Language Model (henceforth, LM), and hypothetical search. The process starts with the capture of an audio signal and then followed by the feature processing and extraction constituents that considered an audio signal as input; mend the speech by eradicating unwanted sound using AM and channel distortion, convert

[2] www.readingrockets.org/article/speech-recognitionlearning accessed 9th May, 2021.

the signals from the time domain to a frequency domain and extract vector features using LM that stand out to the Speech Recognition as result. The diagram below best describes the process:



Source: Architecture of an ASR System-based, Nenny *et al*, (2018)

Diagram 1: Architecture of ASR system

How can Computer Accept Speech as Data?

Speech-activated applications have become part of our daily lives. A sound is an analogue and a computer recognizes only digital inputs. For a computer to accept speech through a microphone, it needs to be transformed to the format recognized by a computer system. A microphone is an input device that is connected or installed in the computer. It is a small magnet covered with a coil of wire; when it vibrates, it creates an electric current in the wire via electric magnetic induction. The amplitude is converted into a voltage which can be read by a computer. From this, frequencies are isolated using Fast Fourier Transformation (henceforth, FFT). FTT is significant in audio and acoustic measurement and it converts a signal into individual spectral components. It is used for fault analysis, quality control and monitoring of systems. With the transformation, the result can be represented by a spectrogram, with time and frequencies labeled as a file.

Every language has a phonetic collection consisting of the sounds that are used in its speech which are the building blocks from which loads can be made. These sounds are called phonemes and spectrogram allows us to identify them. As human speech varies due to accents and variations of language, a Neural Network (henceforth, NN) is introduced as an alternative to machine language learning. NN undergoes a process the way the human brain operates; it is a series of algorithms that identify the core relationship in a set of data. It is capable of improving itself with input data. NN performs change on the data in input and output layers in the interconnected nodes. When sounds are identified, the system begins to analyze words, phrases and sentences. It first, identifies phonemes and processed them through the use of the Language Model (henceforth, LM). It then conducts a syntactic analysis to sort useful and meaningful sentences through investigating word order in accordance with grammar rules using parsing trees and finally bring out the output.

Google Cloud Speech API

Listening and Visual skills are part of language skills, as hearing and seeing are becoming an important part of input activities. Google is building applications using machine learning that uses this form of inputs. Google Translate and Google vision API are examples of such applications. These two APIs are integrated into Google Cloud Speech API, with such a complete API developing an application that can translate, language which becomes easy for the software developers. With the use of Google Cloud, Speech API developers can turn audio into text using Neural Network Models. The Google Cloud Speech API recognizes more than one hundred and ten (110) languages and variants, to support a global user-based, (Nenny, *et al*., 12). Among other functionalities of this API is the ability to dictate using the application microphone, the use of speech to command and control applications, writing of audio files, recognizing the uploaded audio-on-demand, and integrate with the audio stored in the Google Cloud Storage, (Stenman, 15). API detects all incoming data in this case audio signal on cyberspace as it is the messenger that welcomes and coveys the data to its final destination and returns a result(s).

Overview of Some Technical Terms

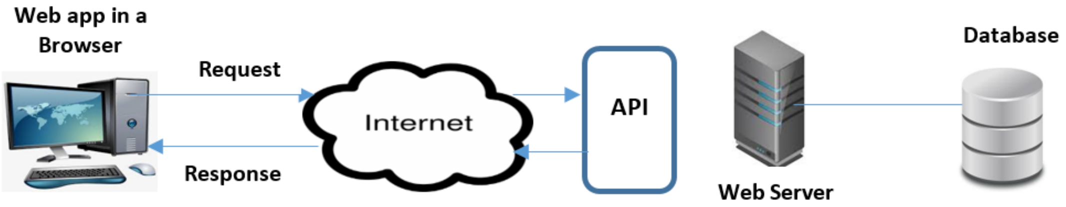
Application Programming Interfaces

How data move from here to there on the cyberspace or Internet?

The answer to the above question is the Application Programming Interface (henceforth, APIs). Connectivity is an incredible thing, as we are connected to the Internet like never before. APIs in a simple term referred to as a go-between that accepts a request, interprets and returns a response. API is like a language two systems used to communicate. An instance is when a user wants to check weather updates, s/he puts the details by identifying his location then send a request and API interferes between user and system by taking the request to the desired destination and providing a result (response).

So, API is a messenger that receives requests from the user and communicates to the system what you want to do and gives you a response. An API consists of two components which comprise technical specification and software interface. In addition, APIs are designed in different forms and can be classified according to the systems that they are designed for. We have the database, operating systems, remote and web APIs. In our study, Google Speech Recognition is an example of a web API. Its function is to standardize data transmission between web services to communicate with each other. As supported by Stylos, *et al*., [9] that "APIs are applications that invoke services or data provided by a software application through a set of existing software resources, such as methods, objects, or Uniform Resources Identifiers" (henceforth, URIs). This summarizes all the complexity behind such resources, as APIs support communication between users/clients and servers. They have consistent communication with clients with URIs guidelines throughout the system, cyberspace.

In modern software design and architecture, the use of APIs delivers "high-level abstractions, simplify and facilitate programming tasks, high-level support in the design of distributed and modular software applications and code reusability", (Robillard, 12). The simplicity makes API flexible to use in speech recognition applications like Google Speech Recognition. Consider the following diagram explains how API works:



Source: How API works. https://www.altexsoft.com/blog/engineering/what-is-api-definition-types-specifications-documentation/

Diagram 2: How API Works

Between software products and contains rules of the data transmission around cyberspace. To have a clear view of how API works, consider API as a waiter in a restaurant, imagine you sitting on a chair and table with a menu to order from. Then the kitchen is part of the system to prepare your order. What is missing here is a link to communicate your order to the kitchen and deliver your food back to your table. That is where the waiter or API comes in. The waiter or API is a messenger that takes a request or order and informs the system in this case, the kitchen what to do and then delivers the response (food) back to you. The same procedure is applied with API on the Internet.

Google Cloud

Google Cloud is also known as Google Cloud Platform, the provider of computational resources for developing, deploying and operations of applications on the internet. "Cloud Platform is a set of modular cloud-based services that provide building blocks you can use to develop everything from simple websites to sophisticated multitier web-based applications". Krishnan, *et al*., [5]. A lot of web applications nowadays explore Google Cloud services.

Voice Recognition

It is the process through which machines detect, analyze and respond to sounds mostly produced by humans is referred to as voice or speech recognition. The history of computerized voice or speech recognition systems can be traced as far back as the early 1930s when Bell Labs began researching for computerized transcription of human speech. Voice recognition became more widespread as personal computers became more widely used, despite a lot of hitches were encountered during the developmental stages of computerized speech recognition systems such as accuracy. It is still considered as a second text input method to the keyboard by a lot of computer users around the globe.

Language educators and call developers started exploring the use of speech recognition systems not long ago and became interested in integrating speech recognition technology with language teaching and call activities, especially with language production practice, (Daniels, 7). He maintained that speech recognition software is being used in language learning. An example is "Subarashii, an interactive dialog system for learning Japanese and in a voice interactive French language training system (ECHOS)".

Voice User Interface

A Voice User Interface (henceforth, VUI) is a subfield of User Interface that comprises of Graphical User Interface (henceforth, GUI) and Command Language Interface (henceforth, CLI). VUI is an intersection space where human and machine occurs with efficient control of human over machine. It is an advanced technology that allows users to interact with machines through the use of voice/speech commands. It means without VUI, it will be difficult to interact with systems or devices. As supported by Sharma, *et al*., (3) VUI is "the interface to any speech application which controlled a machine by simply talking to it". The researchers added that designing a good VUI requires interdisciplinary talents of computer science, linguist and human factors psychology – all of which are skills that are expensive and hard to come by. Even with advancement tools, constructing an effective VUI requires an in-depth understanding of both the task to be performed as well as the target audience that will use the final system.

The VUI is a hands-free technology and is considered to be an Artificial Intelligence that makes it easy, efficient, less time-consuming and a way of interaction. It is a way of interacting with machines or systems using a human's voice rather than typing text. VUI is an interface that enables users to use voice input to control computers, systems, or devices. Functional VUIs depend on speech recognition to translate voice input into commands that a machine can understand. VUI uses speech recognition tools to convert the voice command into text and the process is the same as that of GUI. Some examples are Google Voice Search and Google Assistant.

Researchers conducted by Schlkwyk, *et al*. (2010) and Liakin, *et al*., (2015) [8] on speech recognition using an interface. Schlkwyk, *et al*. (2010) researched Speech Recognition using Multimodal User Interface in Google Mobile App (GMA). For each key area of acoustic modeling and language modeling, the researchers describe some of the challenges faced as well as some of the solutions developed to address those unique challenges. The researchers review some of the common metrics to evaluate the quality of the recognizer. They describe the algorithms and technologies used to build the recognizer for Google search by speech.

Also, Liakin, *et al*., [5] investigate the second language acquisition of the French vowel /y/ using a mobile-assisted learning environment, via the use of automatic ASR. An experimental study was conducted with French learners assigned in groups for some time to see the result of their pronunciation activities that include the use of the application. The study applied a pretest/posttest design. According to the results of the dependent samples t-tests, only the ASR group improved significantly from pretest to posttest (p < 0.001). The ASR production measures suggest that this type of learning environment is propitious for the development of segmental features such as /y/ in French.

Findings and Challenges of Google Speech Recognition

To this effect, in Usmanu Danfodiyo University, Sokoto, there is a current blending or transformation from normal teaching-learning to the virtual teaching-learning process to avoid the hitch caused by the occurrences of the Covid-19 pandemic in the future. Ashwell, *et al*., [7] encourage that the accuracy in Google API's native languages is still a challenge "despite its overall 89.4% accuracy score". As opposed by Matarneh, *et al*., [20] that "even though Google Speech Recognition has been reported as the most convenient API as the result of its computing power", there are some challenges in both its technology and adaptations in the Nigerian context. For them, it is also important to know that "any pronunciation that is difficult to French speakers might be equally difficult for the Google Speech API to recognize it". This shows that pronunciation is a key factor for speech recognition, processing using acoustic filtering and the capture as input to the system to transcribe it by allowing a voice to serve as an interface between users (humans) and computer systems also enable data transmission between one software product and another.

Users that benefit from speech recognition technologies for learning include but are not restricted to users with learning disabilities, poor or limited motor skills, vision impairment and limited language. Lack of knowledge is a challenge despite its flexibility and convenience. It rises contact with a computer, improvement in writing mechanics, increases independence, decreases nervousness around writing and improvements in basic reading and writing abilities.

Conclusion

Technology touches all parts of life. Google provides different applications including APIs which are web-based. Due to this fact, we overviewed some terminologies in understanding the process and concept of Google Speech Recognition API such as Google Cloud for storage, web-based software. We found the relevance of Google Speech Recognition in teaching-learning the French language. We overviewed some concepts of Google Speech Recognition and define Speech recognition which is also referred to as voice recognition in some cases.

The French language is recognized in globalization as we identified its status and position internationally in terms of usage on the internet and technology, learners as well as speakers. The research showed that good pronunciation helps speech recognition capture good data which in return helps in the stages involved before the computer accepts it and determines what is being said. The stages start with a microphone that receives and converts the speech through a process into digital input. Then, the computer analyzes it using NN, as words and phrases are being picked up by analyzing that sound until inputs are determined and finally the computer acts. Some of the benefits of speech recognition technologies for teaching-learning helps users with learning disabilities, poor cognitive skills, inadequate French language and vision deficiency.

We answered the question of how data move from one place to another. This is a result of API which is an interface between humans and systems. API serves as a messenger in a company that moves files from one office to another and retrieved them after approval for onward communication to the applicants. We, therefore, gathered some findings as well as challenges to the application. Some challenges found were inadequate knowledge of API even though it is flexible and convenient. It increased contact with computer and independence, improve writing mechanics, it decreased anxiety around writing and improve competence in reading and writing abilities. Some diagrams were drawn for a clear understanding of the process of using API in language learning as well as how API itself works. Some instances were also given to understand the role of API in Google on the Internet.

Future Work

As Usmanu Danfodiyo University, Sokoto is on top gear in embracing the current trend of transformation from normal teaching-learning to the online teaching-learning process, we intend to propose and develop a web application which we will integrate Google Speech Recognition API to harness the most recent advancement in Speech Recognition technology to aid the language teaching-learning process of the institution.

References

1. Adeyinka-Ojo S, *et al*. Covid-19 Pandemic and adoption of Digital Technology in Learning and Teaching. https://doi.org/10.15224/978-1-63248-190-0-03, 2020.
2. Anggraini N, *et al*., Speech Recognition Application for the Speech Impaired using the Android-Based Google Cloud Speech API. Telkomnika (Telecommunication Computing Electronics and Control), 16(6), 2733-2739, https://10.12928/Telkomnika.v16i6.9638, 2018.
3. Ashwell T, *et al*. How accurately can the Google Web Speech API Recognize and Transcribe Japanese L2 English Learners' Oral Production?. 2018; 13(1):59-76.
4. Daniels P, *et al*. The Suitability of Cloud-Based Speech Recognition Engines for Language Learning. JALT CALL Journal. 2017; 13(3):229-239.
5. Hincks R. Speech Recognition for Language Teaching and Evaluating: A study of Existing Commercial Products. In Seventh International Conference on Spoken Language Processing, 2002.
6. Krishnan SPT, *et al*. Building Your Next Big Thing with Google Cloud Platform. https://doi.org/10.1007/978-1-4842-1004-8, 2015.
7. Liao PA, *et al*. What are the Determinants of Rural-Urban Digital Inequality among School children in Taiwan? Insight from Blinder-Oaxaca Decomposition. Computers and Education. 2016; 95:123-133.

8. Liakin D, *et al*. Learning L2 Pronunciation with a Mobile Speech Recognizer: French /y/.CALICO Vol. 32.1 2015 1–25 Equinox Publishing, 2015.
9. Matarneh R, *et al*. Speech Recognition Systems: A Comparative Review, 19(5):71-79. https://doi.org/10.9790/0661-1905047179, 2017.
10. Meng M, *et al*. Application Programming Interface Documentation: What do Software Developers Want? Journal of Technical Writing and Communication, 48, 295-330. https://doi.org/10.1177/0047281617721853, 2018.
11. Myers BA, *et al*. Improving API Usability. Communications of the ACM. 2016; 59(6):62-69.
12. Robillard MP. What Makes APIs Hard to Learn? Answers from developers. IEEE Software. 2009; 26(6):27-34.
13. Schalkwyk J, *et al*. Google Search by Voice: A case study. Google, Inc. Mountain View: Canada, 2010.
14. Sharma K, *et al*. Exploring of Speech Enabled Systems for English. Department of Computer Science Applications, Teerthanker Mahaveer University, International Conference on System Modeling & Advancement in Research Trends (SMART), 2012.
15. Stenman M. Automatic Speech Recognition, An Evaluation Of Google Speech. UMEA University, 2015.
16. Stylos J. Making APIs More Usable With Improved API Design, Documentation And Tools (Doctoral dissertation). Carnegie Mellon University. Retrieved from http://www.cs.cmu.edu/_NatProg/papers/, 2009.
17. Stylos J, *et al*. Improving API documentation Using API Usage Information. In Proceedings of IEEE Symposium on Visual Languages and Human-Centric Computing (pp. 119–126). Washington, DC: IEEE, 2009.
18. Zaki MZ, *et al*. Translation and Modern Technologies: An Appraisal of Some Machine Translation. Degel: Journal of Faculty of Arts and Islamic Studies, Vol. 15, December, ISSN 0794 9316, 2017.
19. Zaki MZ, *et al*. The Importance of Information and Communication Technology (ICT) in the learning of French Language in Nigeria. JALAL: Journal of Languages and Literature. Department of Modern European Languages and Linguistics, Usmanu Danfodiyo University, Sokoto-Nigeria. 2015; 6(2):163.
20. Altex Soft. What is API: Definition, Types, Specifications, documentation?, 2019. https://www.altexsoft.com/blog/engineering/what-is-api-definition-types-specifications-documentation/
21. Mule Soft Videos. What is an API? https://www.youtube.com/watch?v=s7wmiS2mSXY 20 June, 2015.