A REPORT ON
Multimedia-based English Vocabulary and Handwriting Tutoring System

By

Harshan B 2018A7PS0166P
Navaneeth Raghunath 2018A7PS0185P
Ratnaraj Wora 2018A4PS0297P

Prepared in partial fulfillment of the Course - CS F407
(Artificial Intelligence)
At

BIRLA INSTITUTE OF TECHNOLOGY AND SCIENCE, Pilani
(Rajasthan) November 2021

Problem Statement

In today's world, English is fast becoming the de facto modicum of communication between people in various parts of the world. Being able to read and write in English is a skill that can fast-track careers and provide exposure to many more opportunities to a person. There are many people in our country who have not been exposed to a good quality of English education due to lack of resources and are consequently denied access to such opportunities. We believe that it is possible to help such people by leveraging the power of Artificial Intelligence. Our goal is to reach the sections of society which stand to benefit significantly from exposure to the English language, including children and adults who are new learners.

Intelligent Tutoring Systems (ITSs) are computer systems that aim to provide personalised instruction and feedback to users, often through AI technology and without a human teacher. As part of our project, we have proposed a novel ITS that will provide an immersive learning experience for those who are learning English for the first time.

Several studies have shown that learners are able to grasp and retain information much more effectively when they learn from diverse sensory cues such as images and through writing. Thus, the best way to reach new learners of the language would be to teach them simple English words with their pictures, and also how to write down the words correctly. This is exactly the target of our focus as part of the ITS we have designed.

Implementation details

The technology stack comprises of Python primarily for the Learner, Expert and Pedagogy components, with HTML with React Javascript framework for User Interface. Handwriting Recognition component has been made using PyTorch deep learning framework for Python. The overall app is made up of three main blocks, the Frontend component, Backend component and the Deep Learning recognition system.

The frontend has a main menu where the learner is supposed to enter one's name. After pressing the 'Start Quiz' button, the learner is taken to a Question Screen where an image is shown. The learner is expected to write down the word legibly on a piece of paper or whiteboard using a pen or a marker. The learner then takes a photo of the written word and uploads it on the Question screen by clicking on the "Browse" button, and clicks on "Upload". The learner is taken to the Answer Screen where the written word and the correct answer are displayed. This repeats for all the questions in the question bank. At the end of the session, the final score is displayed.

The photo of the written word is sent in JSON format to a REST API endpoint which performs the recognition and returns the recognized written word in text form. This endpoint is implemented as a Flask application, which is a Python webserver. The Flask app contains the Text Recognition module which in turn consists of two parts: Character Segmentation and Character Recognition submodules.

The Character Segmentation submodule performs a variety of pre-processing operations on the image like Gray-Scaling, Eroding, Dilating, Thresholding, etc. This pre-processed image is split into various segments using a Canny Edge filter. Each of these segments are then cropped and are individually sent to the Character Recognition submodule which uses a Deep Learning-based PyTorch model trained on EMNIST dataset. The predicted characters are concatenated into a word and are returned as a JSON response to the REST API request.

Block diagram
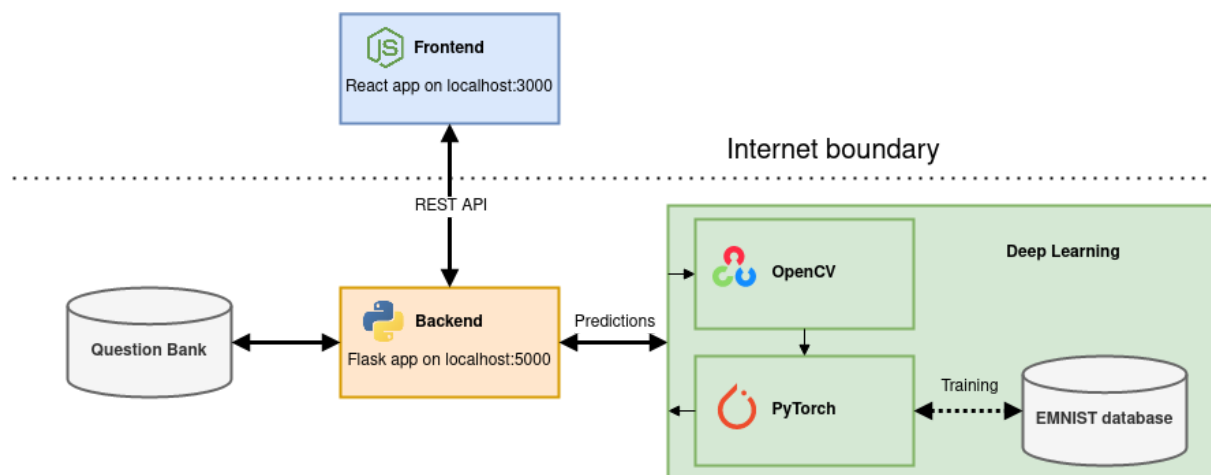
How to execute

- <u>Hardware Requirements:</u> Nvidia GPU with CUDA support.
- Install Python>=3.6 and NodeJS>=14.18.1.
- Install required Python and NodeJS dependencies:
  - Python: `pip install -r requirements.txt` in `backend_dl` directory
  - NodeJS: `npm install` in `frontend` directory
- Run `python app.py` in `backend_dl` directory and `npm start` in `frontend` directory, in separate terminals.
- The Backend module (app.py) is a Flask app which runs on `localhost:5000`.
- The Frontend module (NodeJS) is a React app which runs on `localhost:3000`.
- Open the proposed ITS application on `localhost:3000` in a browser.

Comparison with existing tools and software

All previous works in this domain have been geared towards those learners who already possess a fundamental understanding of English and are looking to improve their proficiency.

The most pertinent is the work by Xu et al.[1], which lays emphasis on teaching non-native speakers to ask for and provide directions in English to different landmarks; this still relies on existing knowledge of basic English comprehension and literacy.

A study conducted by the paper[3] makes use of a virtual tutor and a voice recognition system to improve the vocabulary and language proficiency of Latinx second grade English learners (ELs). The virtual tutor provides lesson content to the students through carefully scaffolded activities. In each of these activities, students were asked to provide their answers along with an explanation to the virtual tutor via a microphone. These answers are then analysed by the voice recognition software to improve the accuracy of the reactions of the virtual tutor. However, there is no written component to the learning, nor does it make use of pictorial cues to stimulate retention in early learners.

At present, there are no systems that focus on those who are completely new to the language. Our solution bridges this gap and allows new learners to acquire the fundamentals necessary to take advantage of such existing methods. It also takes advantage of deep learning to provide an interactive and involving learning experience as part of the Intelligent Tutoring System.

Analysis and Conclusions

As mentioned above, it can be demonstrably shown that learners show a much faster rate of understanding and recollection when presented with a variety of sensory information, such as pictures, multimedia etc. due to the link that the brain generates between the word and its image. Furthermore, learning by writing is one of the age-old techniques that has been shown to produce great improvements in the retention and recall of students.

For early learners of a language, taking advantage of both of these methods is a sure-shot way to make the lessons engaging, and ensure that they retain the information that they are being taught. This is of vital importance especially in teaching a new language such as English, to the target demographic of underprivileged and marginalised sections of our society, and can prove invaluable in boosting their involvement and learning curve.

Our application takes these factors into account while introducing learners to the language. They are presented with pictorial cues and are made to write down the word that it represents. The questions are carefully chosen to be simple enough to not discourage new learners while also covering the most commonly used words by new learners. We expect the learning curve will be a steady and smooth rise to reading and writing proficiency with a very low barrier to entry to make it as accessible as possible.

Future Improvements

Future iterations of the application can incorporate more useful features such as a difficulty system that would dynamically present learners with more or less difficult questions based on their current relative proficiency and skill level. We could also make use of their answer history to target and work on improving weak areas for each specific user. Audio and video

cues can be made use of alongside imagery to make the learning experience even more holistic. Such improvements would make it an invaluable system for introducing learners to the English language.

Team

Harshan B	2018A7PS0166P
Navaneeth Raghunath	2018A7PS0185P
Ratnaraj Wora	2018A4PS0297P