International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 06 | June 2020 www.irjet.net p-ISSN: 2395-0072

GENERATING AND MODIFYING IMAGES BASED ON LINGUISTIC INSTRUCTION

Anurag M Khot [1], Ahmed Sarfaraz J[2], Anush H[3], Chethan V[4], Manjunath S[5]

[1,2,3,4]Students, Department of Information Science & Engineering, Global Academy of Technology, Bengaluru, Karnataka, India
[5]Assoc. Professor, Department of Information Science & Engineering, Global Academy of Technology, Bengaluru, Karnataka, India

---

Abstract - *Photos and videos appeal to our brain. High quality visual content have a major impact. Visual content communicates more effectively and efficiently. The ability to produce image based on voice input is very interesting and provides to be very useful. It has various applications where one of them is for educational purposes where visual learning can be very effective because visuals are processed 60,000 times faster than text and retains longer in memory than text. This is one of the many applications of the technology. In this motive, a method is proposed to develop a speech-controlled text to image conversion where the user provides input through speech. It is important to take necessary measures with the technology and to help them to live with the current world irrespective of their vision. In the motive of helping them, a method is proposed to develop a voice-controlled text to speech module in order to read, store and understand the text in an easier and faster way. It is advantageous for the people to understand the concepts more efficiently. A required dataset is loaded in order to match the text present with the given input, once keywords are matched, it is synthesized for producing appropriate image as output.*

Key Words: Python Programming, Speech to text, Identification of nouns and ad-positions, Sketch Output, Data Storage.

1. INTRODUCTION

In everyday life, speech is considered as one the most important medium of communication. While conveying the message the most widely used form can be termed to be a speech signal. When a user speaks to a conversational interface, the system has to be able to recognize what was said. The speech-to-text component processes the acoustic signal that represents the spoken utterance and outputs a sequence of word hypotheses, thus transforming the speech into text. The speech-to-text refers the conversion of speech signal to the text format. Further analyse the converted text to extract the nouns and ad positions using Natural Language Processing (NLP). The Quick Draw dataset is a collection of 50 million drawings over 345 categories. Recognize Stream, get Entities and analyze Syntax are used to realize the input part. Identified entities are displayed on screen.

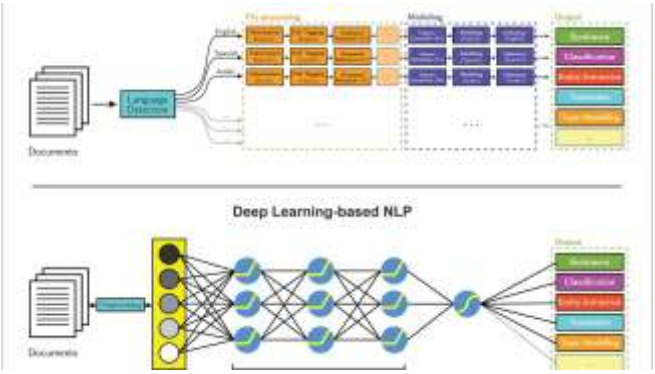

Fig 1: Deep learning based NLP

2. OBJECTIVES

We live in the 21st century right now. From cartwheels to self-driving cars, from Flyer 1903 to SpaceX landing rockets, from tin can telephones to smart phones, humans have come a long way. But there is no much change in the form of education delivered to students. Knowledge and education leads to more innovative ideas. The brain remembers images better compared to words because it is easier to understand that way. Words and images go side by side to understand a concept. The same can be implemented in classroom teaching where it helps a child understand better if the teachings were in a pictorial form. To keep their imagination alive and also help them understand better. We are introducing a system where speech is taken as input and this speech is converted to a text and then to sketch later. The words used by the teacher form a sketch as they speak which keeps the class more dynamic. Students get a better idea of the concept. This also helps them get more ideas because reality still leaves a lot to imagination.

3. DEVELOPED METHODOLOGY

A. Speech-To-Text

Speech is recognized through audio input and SpeechRecognition is used to make this retrieval of input really easy. SpeechRecognition is used because writing python programs for accessing and processing the audio files from scratch would take a deliberate amount of time, instead using this will have us ready and good to go in a few minutes. The SpeechRecognition library is extremely

flexible because it acts as a wrapper for several popular speech APIs. One of these is the Google Web Speech API which supports a default API key that is hard-coded into the Speech Recognition library. That means you can get off your feet without having to sign up for a service. Speech Recognition package is an excellent choice for any Python project because of its flexibility and ease of use. However, support is not guaranteed for every feature of each API it wraps. Much time will be needed to be spent in researching the available options to find out if Speech Recognition is the best choice to work with in your particular case.

B. Extracting the nouns and ad positions using Natural Language Processing (NLP)

The Natural Language Toolkit (NLTK) is a set of libraries used for English written in the python programming language for symbolic and statistical natural language processing (NLP). Tokenization, parsing, classification, stemming, tagging and semantic reasoning are text processing libraries housed within the library. Graphical demonstrations and sample data sets are included as well as a cookbook and a book to explain the principles behind the underlying language processing tasks that NLTK supports. To predict how people are feeling the use of Language, tone, and sentence structure can explain a lot, and a combination of the Natural Language Toolkit can even be used to predict how people might feel about similar topics, a python library used for analyzing text, and machine learning. For understanding the concepts involved, a single body of text to clean and analyze key parts of past presidents' inaugural speeches, which are included in NLTK's corpus library. Once there is a strong foundation of the basics, applying these techniques to a machine learning classification would be a breeze of a task to accomplish and would be able do with just about any text-rich data.

C. Displaying related sketch

The Quick Draw Dataset which was contributed by players of the game Quick, Draw. It consists of a collection of drawings which is accountable for millions across various categories, metadata is included which consists what the player was asked to draw and in which country the player was located and is tagged and these drawings were captured as timestamped vectors with those tags. Simple Dataset is as follows:

{ "key_id":"5891796615823360",

"word":"nose",

"countrycode":"AE",

"timestamp":"2017-03-01 20:41:36.70725 UTC",

"recognized":true,

"drawing":[[[129,128,129,129,130,130,131,132,132,133,133,133,133,...]]] }

NDJSON is a convenient format for storing or streaming structured data that may be processed one record at a time. It works considerably well with Unix-style text processing tools and shell pipelines. It's a great format for log files. It's considered a flexible format for passing messages between cooperating processes. Identified entities are matched with datasets; if found sketch is displayed else wait for the new input.

D. System Architecture

System architecture is the conceptual model that defines the structure, behaviour, and more views of a system. An architecture description is a formal description and representation of a system, organized in a way that supports reasoning about the structures and behaviours of the system.

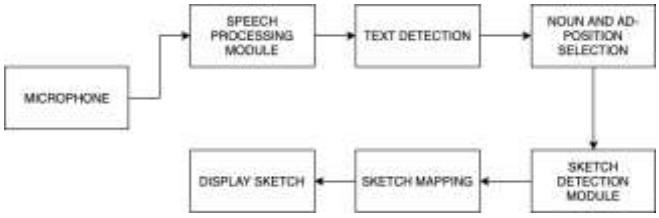

Fig. 2 System Architecture

Step 1: With Speech Recoginition module and NLP module, nouns and adpostions are extracted like "above", "under" and so on from the real time streaming.
Step 2: Recognize Stream, get Entities and analyze Syntax are used to realize the input part.
Step 3: Identified nouns are matched with datasets; if found sketch is displayed else wait for the new input.

E. Low Level Design

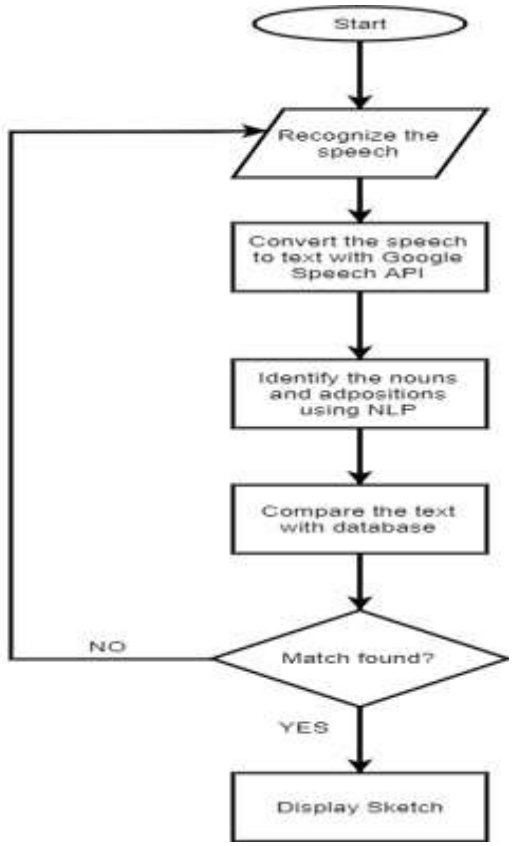

Fig. 3 Flow of processes

With Speech Recognition module and NLP module, nouns and adpostions are extracted like "above", "under" and so on from the real time streaming.

Recognize Stream, get Entities and analyse Syntax are used to realize the input part.

Identified nouns are matched with datasets; if found sketch is displayed else wait for the new input.

Finally, a recurrent image generation model which takes into account both the generated output up to the current step as well as all past instructions for generation. The model is able to generate the background, add new objects, and apply simple transformations to existing objects.

F. Tools, Libraries and Datasets

The application is built upon various libraries using various tools and datasets. All of them have a crucial part to play in building the complete application with the optional interfaces.

Some of the libraries, tools and datasets with its uses are:

1) NLTK**:** The Natural Language Toolkit (NLTK) is a platform used for building Python programs that work with human language data for applying in statistical natural language processing (NLP).
It contains text processing libraries for tokenization, parsing, classification, stemming, tagging and semantic reasoning. It also includes graphical demonstrations and sample data sets as well as accompanied by a cook book and a book which explains the principles behind the underlying language processing tasks that NLTK supports.
2) VS Code: The Visual Studio integrated development environment is a creative launching pad that you can use to edit, debug, and build code, and then publish an application.
3) Python: It is an interpreted, high-level, general-purpose programming language. It is dynamically typed and garbage-collected. It supports multiple programming paradigms, including structured object-oriented, and functional programming. It has a comprehensive standard library. The application is built using this programming language.
4) Java: The primary objective of Java programming language creation was to make it portable, simple and secure programming language. Apart from this, there are also some excellent features which play an important role in the popularity of this language.
5) OS MODULE: The OS module in python provides functions for interacting with the operating system. OS, comes under Python's standard utility modules. This module provides a portable way of using operating system dependent functionality. The *os* and *os.path* modules include many functions to interact with the file system. It is possible to automatically perform many operating system tasks.
6) PYAutoGUI: PyAutoGUI lets your Python scripts control the mouse and keyboard to automate interactions with other applications. The API is designed to be as simple. PyAutoGUI works on Windows, macOS, and Linux, and runs on Python 2 and 3.

4. EXPERIMENTAL RESULTS

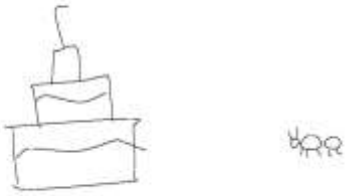

Fig 4 : THE ANT WAS HEADING TOWARDS THE CAKE

Fig.5: THE MAN WAS DANCING IN THE RAIN

Fig 6: THE CARROT WAS WITH RABBIT

5. TESTING

Table I shows the set of test cases which were tested on the application and yielded the following results in the interface designed.

Input (Microphone)	Expected Result	Actual Result	Correctness
1. Radio on the Table	Radio on the Table	Radio on the Table	Yes
2. Man was dancing in the Rain	Man was dancing in the Rain	Man was dancing in the Rain	Yes
3. Carrot was with Rabbit	Carrot was with Rabbit	Carrot was with Rabbit	Yes
4. Monkey on the Tree	Monkey on the Tree	Monkey on the Tree	Yes
5. Hat below the Table	Hat below the Table	Heart below the Table	No
6. Man waiting outside door	Man waiting outside door	Man waiting outside door	Yes
7. Radio below the Table	Radio below the Table	Radio below the Table	Yes
8. Ant was towards the Cake	Ant was towards the Cake	Ant was towards the Cake	Yes
9. Boat was sailing in the Water	Boat was sailing in the Water	Boat was sailing in the Water	Yes
10. Clock hanging on the Wall	Clock hanging on the Wall	Clock hanging on the Wall	Yes

Table- I: Test cases for sentences

6. CONCLUSION

We take in speech as input and convert it into a sketch. We would like to introduce the idea of interactive classroom session where the teacher explains the concepts through sketches for ease of understanding and keep their imagination alive. The words used by the teacher are converted to text using speech recognition API and then these texts are converted to a related sketch using NLP. The images are created dynamically as the words are spoken which reduces the pre-class preparation for the teachers and makes the class more interactive. This concept can be developed further as the limit is infinity and beyond.

We reviewed important prior works about text-to-picture systems and approaches. We compared them according to input knowledge resolution, knowledge resources, knowledge extraction, image selection, and matching and output rendering. For the most part, existing systems favour linguistic-focused text processing. The objective of this

review is to investigate the feasibility of automatic visualization of text through multimedia, which, in turn, involves text processing and analysis with the available tools. Moreover, this review allows us to refine our problem statement and to identify relevant topics for further research. So far, we propose the following approach: First, processing the story to get a semantic representation of the main characters and events in each paragraph. Then, constructing expanded queries for each paragraph using the output of the previous step. Third, through an image search, finding a list of the top picture candidates. Exploring the results, a user or instructor can eventually refine the results of the automatic illustration step by selecting a few suitable pictures to compose the final visualization for each paragraph.

ACKNOWLEDGMENT

We are grateful to our Institution, Global Academy of Technology, with its ideals and inspirations for bringing in the quality in the project work carried out at this institute.

We earnestly thank our Principal, Dr. N. Ranapratap Reddy, and our HOD, Dr. Ganga Holi, Global Academy of Technology for facilitating a congenial academic environment in the College and for their kind support, guidance and motivation during the course of our project work. We would like to extend our sincere thanks to our parents and friends for their support.

REFERENCES

[1] Alaaeldin El-Nouby, Shikhar Sharma, Hannes Schulz, Devon Hjelm, Layla El Asri, Samira Ebrahimi Kahou, Yoshua Bengio, Graham W. Taylor" Tell, Draw and Repeat: "Generating and Modifying Images Based on Continual Linguistic Instruction".

[2] William Chan, Navdeep Jaitly, Quoc V. Le, Oriol Vinyals"Listen, Attend and Spell "in August 2015.

[3] Oriol Vinyals, Alexander Toshev, Samy Bengio, Dumitru Erhan. "Show and Tell: A Neural Image Caption Generator"-in November 2014.

[4] Marcia A. Bush,"Speech and Text-Image Processing in Documents"-HLT,1993.

[5] Golan Levin, Zachary Lieberman, "In-Situ Speech Visualization in Real-Time Interactive Installation and Performance"-in 2004.

[6] Shikhar Sharma, Dendi Suhubdy, Vincent Michalski, Samira Ebrahimi Kahou, Yoshua Bengio,"ChatPainter: Improving Text to Image Generation using Dialogue"-in Feb 2014.

[7] Tingting Qiao, Jing Zhang, Duanqing Xu2, and Dacheng Tao. "MirrorGAN: Learning Text-to-image Generation by Redescription"-in Mar 2019.

[8] Sharon Gannot, Marc Moonen, "The application of the unscented kalman filter to speech Processing"-in Sep 2003.

[9] Elizabeth Coates, Andrew Coates, "The essential role of scribbling in the imaginative and cognitive development of young children"-in Apr 2015.

[10] Scott Reed, Zeynep Akata, Xinchen Yan, Lajanugen Logeswaran Bernt Schiele, Honglak Lee, "Generative Adversarial Text to Image Synthesis"-in May 2016.

BIOGRAPHIES

Anurag M Khot is currently pursuing Bachelor of Engineering in Information Science at Global Academy of Technology which is affiliated to Visvesvaraya Technological University of Belagavi, India. Mainly interested in development and programming. Undertaken few mini projects and paper presentations.

Ahmed Sarfaraz J is currently pursuing Bachelor of Engineering in Information Science at Global Academy of Technology which is affiliated to Visvesvaraya Technological University of Belagavi, India. Mainly interested in development and programming. Undertaken few mini projects and paper presentations.

Anush H is currently pursuing Bachelor of Engineering in Information Science at Global Academy of Technology which is affiliated to Visvesvaraya Technological University of Belagavi, India. Mainly interested in development and programming. Undertaken few mini projects and paper presentations.

Chethan V is currently pursuing Bachelor of Engineering in Information Science at Global Academy of Technology which is affiliated to Visvesvaraya Technological University of Belagavi, India. Mainly interested in development and programming. Undertaken few mini projects and paper presentations.

Manjunath S is a research scholar at JSSATE Research Centre, Dept. of CSE, JSSATE, affiliated toVTU Belagavi. He completed M.Tech Computer Science and Engineering from NMAMIT Nitte Mangalore affiliated to VTU Belagavi, Karnataka. Now he is working as a Associate Professor Dept. of ISE, Global Academy of Technology, Bengaluru.