EE/CprE/SE 491 BIWEEKLY REPORT 2

September 13 - September 27

Group number: sddec21-06

Project title: DigiClips Media Design

Client: DigiClips

Advisor: Ashfaq Khokhar

Team Members/Role:
Sam Massey - Assignment planning, Research, Work on speech-to-text
Tyler Johnson - Planning and implementing test cases
Maxwell Wilson - Primary point of contact with client, Research, Work on speech-to-text, Team Leader
Max Van De Wille - Documenting architecture changes, Research, Work on video-to-text

o **Weekly Summary**

This past week, our team continued development of speech-to-text and video-to-text elements. While using Deepspeech open API for speech-to-text, we have made progress to allow recordings to be broken up into multiple segments, transcribed, then concatenated back to give a final result of a recording converted to a transcript. On the video-to-text side, we have seen higher accuracy by using different pre-processing steps before passing the image onto Google's Tesseract API for optical character recognition. The video-to-text app is also undergoing a refactoring effort to clean up some of the code and alter its form into a hosted API.

o **Past week accomplishments**

Max Wilson:
- Worked with speech-to-text application to improve efficiency
- Dockerized the fastapi application so that we can easily run the application in a docker container on any system with Docker
- Experimented with AWS to see if our app could be run on a system with better specs. I'm still having trouble getting it running in an AWS environment but it should be close

- Started looking into possible audio-enhancement or preprocessing we can do to improve accuracy

Sam Massey:
- Research how, and increase speed of speech-to-text element.
- Experiment with PyDub to try and increase speech of transcripts created.
- Try increasing audio recording volume to see if speech-to-text will be more accurate.
- Research how to use multithreading to increase speed of program.

Max Van De Wille:
- Tested custom training tesseract OCR with samples of news footage font
- Began refactoring video-to-text script into actual API that accepts filename
- Began working on duplicate filtering for multiple instances of text within a segment
- Tested image pre-processing methods to find most accurate approach

Tyler Johnson:
- Collected representative data from range of available clips of audio and video
- Transcribed portions of audio into text
- Began planning methodology for testing video to text program

o **Pending issues**

- No unified/standardized testing set to compare performance of one iteration to the next makes it hard to benchmark progress/performance improvements.
- One of the main pending issues our team as well as the team at DigiClips is facing will be ensuring accuracy in our speech-to-text and video-to-text.
- Speech-to-text app is currently splitting the audio file without overlap. This could split in the middle of a word and could cause data to be lost. We need to implement overlap in the chunking logic so that we don't cut any words in half.
- Certain fonts displayed in sample videos are not detected as well by tesseract

o **<u>Individual contributions</u>**

Team Member	Contribution	Weekly Hours	Total Hours
Sam Massey	Deepspeech work, PyDub research and experimentation	7	26
Tyler Johnson	Translation, testbench creation	6	24
Maxwell Wilson	AWS Testing, presentation	7	28
Max Van De Wille	Video-to-text development, generating benchmark samples for client	7	26

o **<u>Plans for the upcoming week</u>**

Max Wilson:
- Implement chunking overlap to catch any words that may have been cut in half
- Implement some sort of multithreading system to improve processing speed
  - Current app processes each chunk sequentially. Could use multithreading to improve performance in these sections
- Design method of tagging audio chunks so that Digiclips can see where certain words occurred in the audio file

Sam Massey:
- Continue development of speech-to-text software
- Improve speed of speech-to-text program
- Look into PyDub to see about the possibility of reducing background noise
- Check the accuracy when using PyDub to increase the volume of the recording file

Tyler Johnson:
- Transcribe audio in clips into text based off of human interpretation
- Continue transcribing audio into text
- Determine methodology for testing video to text and what factors to consider

Max Van de Wille
- Continue refactoring video processing script into API
- Refactor utility methods into separate utils file/folder to clean up main running process
- Continue work on duplicate filtering
- Alter output method from one long string to timestamped indices of written text