

Team Name: sddec21-06

Team Members: Sam Massey, Tyler Johnson, Max Van de Wille, Maxwell Wilson

Report Period: Mar 29-Apr 12

Summary of Progress in this Period

In this period, we did a lot of research for different speech-to-text options using different languages and to see what will be the most effective. After our research, our team, as well as the team over at DigiClips, decided to move forward with DeepSpeech open source software using python. In this period, we have also started creating different ways to test the DeepSpeech speech-to-text with the recordings that DigiClips captures. The final, main thing that we have started working on this reporting period is a way to add the functionality for different punctuation checks, such as adding commas or periods in the output of the speech-to-text. We have yet to decide if this is a task that we should create, or if it is a task that can be done with open source software or by using the captions that are also recorded already in the DigiClips database.

Pending Issues

In the coming weeks we will have a couple issues to solve. First, we will need to check with Bob and Henry to get a better idea of how quick the speech to text program should run. We also need to focus on planning how our grammar checker for speech-to-text will work and what kind of capabilities it will have. Before we solidify that plan we will need to talk with Bob and Henry about their expectations for the grammar checker and make sure that their expectations are reasonable. Grammar checking the speech-to-text output is a tricky issue that we will have to focus on in the coming weeks.

Plans for Upcoming Reporting Period

We will be continuing to familiarize ourselves with the DeepSpeech API for speech-to-text and begin building a working pipeline for processing video files, testing using different parameters and grammar/speech checking tools to process the output into proper sentences. For video-to-text, we will be continuing to test some tools for optical character recognition and image processing to determine the best combination for our project. We will also be looking at examples of news videos similar to those that the application will be processing to determine how much of the frame is necessary to capture and process, since in many cases the news video only has text at the bottom.
