

Introduction

This is a short documentation for GenerateLab.py, whose purpose is to create .lab files from a .docx file and .wav files for alignment. I've written it as a Research Assistant under Dr Fatima Hamlaoui at the University of Toronto from July to August 2018.

The code can be found at: <https://github.com/Dragonic/generate-lab-files-for-alignment>

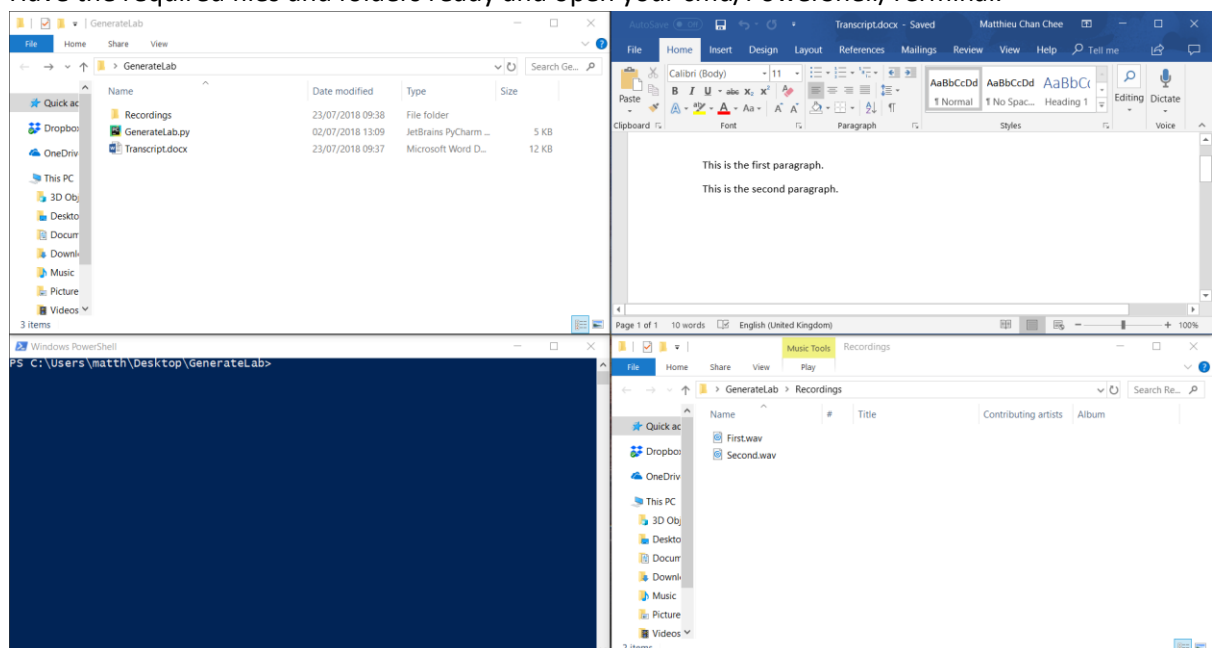
Should you need any clarification, you can contact me at: matthieuuchancee@gmail.com

Requirements

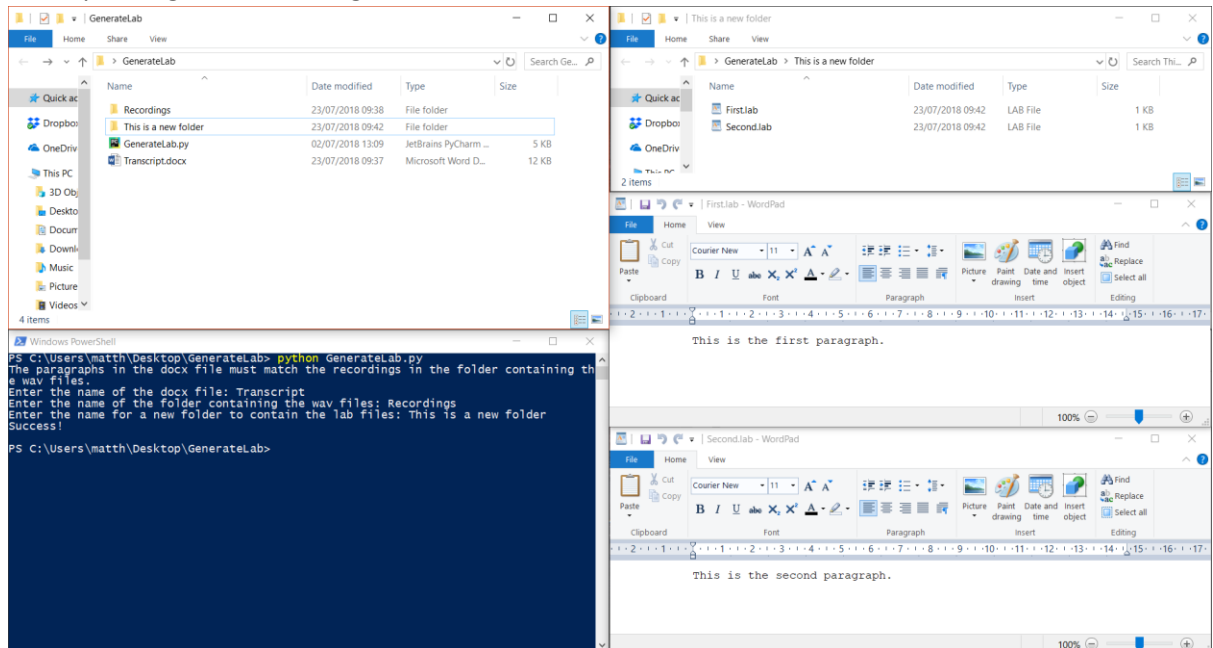
- The .docx file, the folder containing the .wav files, and GenerateLab.py must be in the same folder.
- The paragraphs in the .docx file must correspond to the .wav files.
- You must have Python 3 installed.

Use

1. Have the required files and folders ready and open your cmd/PowerShell/Terminal.



- Run GenerateLab.py and follow the instructions. A new folder containing the .lab files corresponding to the paragraphs in your .docx file and bearing the same names as the corresponding .wav recordings should be created.



Notes

In the following cases, an error message will be displayed and no .lab file will be generated:

- The .docx file does not exist.
- The folder containing the .wav files does not exist.
- Any of the following characters was used to name the new folder to contain the .lab files:

$$\backslash, /, :, *, ?, ", <, >, |$$
- The name you entered for the new folder is already used.
- The number of paragraphs in the .docx file is not equal to the number of .wav files.

Some people end their .docx files by pressing several times on the Enter/Return key, thus creating 'empty' paragraphs. This program ignores empty paragraphs I've been able to predict but should other types of empty paragraphs show up, the code will have to be updated. As of 23 July 2018, an empty paragraph is defined as any of:

- Nothing (empty string)
- 1 to 9 spaces (' ', ' ', ..., ' ')