

Subject: Working on Interview Transcriptions? -- Helpful Tool for Processing Zoom Transcripts!

From: Matthew James Jumamoy <mjumamoy@stanford.edu>

Date: 9/7/21, 10:46 PM

To: "urbst_classof2022@lists.stanford.edu" <urbst_classof2022@lists.stanford.edu>

CC: Michael Kahan <mkahan@stanford.edu>, Deland Chan <deland@stanford.edu>, Morgan Speer <mspeer@stanford.edu>

Hi Fellow Urban Studies Seniors,

If you guys are working on interviews for your senior projects, it may seem a little bit cumbersome to create transcriptions so you can code or do easier analysis between your interviews. However, if you are doing Zoom interviews with your participants and you are using the "Record to Cloud" feature, you can get an automated transcription of the interviews in the form of a VTT file.

However, this VTT file is really messy and has a ton of timestamps and things that you as a human don't really need. But, it gets about 80% of it right-- which can be a serious timesaver.

That's where the Python tool webvtt-py comes into play. If you use webvtt-py and import that VTT file and run it through a script, a beautiful new plain text file transcribing your interview pops out. From there, you can import it into Word, create a more formal looking document, and listen to your interview and fix the errors that Zoom transcription makes. Overall, it saves me over an hour per interview to use Zoom transcription + webvtt-py to transcribe interviews.

Additionally, I developed a small tool that can automate this entire process and export out transcripts that can be put into MS Word/Pages/LibreOffice! Plus, you can replace your interviewee's names with pseudonyms if that's the path that you took!

Here are some abridged instructions if you are familiar with using Python. Feel free to ask me if you'd like to use my tool and I'd be happy to help!

1. Open PyCharm or Visual Studio or the Python Interpreter, make sure Python 3.6 or higher is installed.
2. Install webvtt-py via the command:

```
MacOS:  
python3 -m pip install -U webvtt-py
```

```
Windows:  
py -m pip install -U webvtt-py
```

3. Create a new folder and insert the VTT files from your Zoom interviews.
4. Download my "Export to Clean Transcript" Script from Github at <https://github.com/mjumamoy/expcts/blob/main/expcts.py> (Right click on "Raw" and Save file as...) and place it in the same folder as the VTT files for the Zoom interviews.
5. Change directory in your Python Console to the folder that you created with the script and VTT files.
6. You have four options when running my tool:
 - c | exports out just the transcript in words
 - ctab | exports out the transcript with tabs instead of ":" that is used by default by Zoom
 - ctabanon | exports out the transcript with tabs and replaces names

-canon | exports out the transcript and replaces names, uses the default ":" delineator
Example Usage:

```
Format for EXPCTS.PY  
expcts.py -flag input_transcript.vtt output_text.txt "Name of Participant on Zoom" Pseudonym
```

```
Example 1: I want a transcript with tabs.  
expcts.py -ctab transcript_09072021_0820.vtt output1.txt
```

```
Example 2: I want a transcript with tabs and I'm replacing a participant's name from John Doe to James  
expcts.py -ctabanon transcript_09072021_0820.vtt output1.txt "John Doe" James
```

Hopefully this script can help some of you! Feel free to share this info with whoever may be doing interviews over Zoom and needs to fix up and create their own transcript.

Thanks!

Sincerely,
Matthew Jumamoy
(he/him)

Stanford University '23
Candidate for B.A., Urban Studies
+1 (702) 907-7279 | mjumamoy@stanford.edu