

JSON to VTT Generator

- Source code
 - Location: galaxy/tools/vtt_generator/jsonToVtt_adapter.xml calls the VTT generator shell script.
 - This shell script generates the VTT format transcript(<https://w3c.github.io/webvtt/>) from a Standard AMP JSON.
- The tool can be invoked from Galaxy UI as other tools.
 - If the tool is run individually without any other tool, the user needs to use Get Data / Upload from a computer to ingest the input file into Galaxy before running the tool.
 - When ingesting, choose binary (the default) as the file format. The file then will be copied into a designated location in the Galaxy file system.
 - The tool can also be embedded in the workflow and can take its input from a preceding tool.
- The tool parameters have detailed labels and helpful info for the users to understand what each one is for and its valid value set.
 - \$input_seg: the standard AMP JSON format input
 - \$input_stt: this is the optional input containing the speaker diarization info(usually present if the input is produced as an output of Kaldi)
 - \$vtt_output_file: the result is a VTT file with transcript text as per the input JSON.
- The output file
 - The output file created by the VTT generator by forming statements by concatenating the words in the input JSON.
 - The end of a statement is identified by the change of speaker or punctuation occurrence. This is when the statement is written to the output file.
 - The start and end times of statements are also written in the VTT file as per its format.
 - Once a current statement is written, the script starts building a new statement in the same way as above and this process continues until the end of JSON is reached,