

Python Script for Txt to JSON

I needed a python script that creates one JSONL file out of multiple txt files which are in one folder.

I have multiple text files of conversation data in one folder, "Input.

In each text file, a facilitator named "Chedy" have a conversation with an "unknown speaker." They are going back and forth. The word "Chedy" or the word "unknown speaker" marks a block which each person says before it the other person's turn.

Sometimes there is an empty line in one block but the text belongs to the same speaker.

Here are example of such a text file

```
:
Unknown Speaker
I'm sitting on the couch in my living room. It's about this time of day
. Chedy
And she snapped at you. Is it true? Listen carefully . Notice the conversation be there now . She
snapped at you. Listen carefully . Is it true? She snapped at you. Those of you new to The Work .
There are two answers to that there one syllable. It takes stillness to get in touch with yes or no Yes
, it's true or no it's not. She snapped at you. Is it true?

No. So just breaston that be there now in that call. Continue to listen. And then get in touch
with how you reacted. When you had the thought she snapped at me. Notice what happened
notice your emotional. Notice the images of past future that came to your mind Also all
responses to noticing how you react what happened when you had the thought she snapped at
me

Unknown Speaker
I become defensive angry manipulative yeah my jaws tight lots of images from the
past of other situations where I felt like she was angry with me I see the future visiting her
and my granddaughter and thinking oh boy here we go. We've done this 1000 times lots of
resistance fear panic I need to get her happy
```

The purpose of this python script is to create prompt and completion pairs in

this format and save to .jsonl file (ChatGPT model training Ready!):

```
{"prompt": "<prompt text>", "completion": "<ideal generated text>"}
{"prompt": "<prompt text>", "completion": "<ideal generated text>"}
{"prompt": "<prompt text>", "completion": "<ideal generated text>"}
```

All lines even if separated by an empty line after the string "unknown speaker" should be combined in one line in the format "unknown speaker: +Text" and be considered as "prompt text".

All lines, even if separated by an empty line after the string “Chedy,” should be combined in one line in the format “Chedy: +Text” and be considered as “ideal generated text.” The string “unknown user” indicates the start of a new block.

Make sure that the python script handles utf-8.

Don’t add the line “Transcribed by <https://otter.ai>” which is included in the text files to the JSON file:

The prompt and completion pairs need to be exported to a json file named “session_data”

An example how prompt and completion should look like in the json file:

“prompt text”=“Unknown speaker: I'm sitting on the couch in my living room. It's about this time of day.”

“ideal generated text”=“Chedy: And she snapped at you. Is it true? Listen carefully. Notice the conversation be there now. She snapped at you. Listen carefully. Is it true? She snapped at you. Those of you new to The Work. There are two answers to that there one syllable. It takes stillness to get in touch with yes or no Yes, it's true or no it's not. She snapped at you. Is it true? No. So just breaston that be there now in that call. Continue to listen. And then get in touch with how you reacted. When you had the thought she snapped at me. Notice what happened notice your emotional. Notice the images of past future that came to your mind Also all responses to noticing how you react what happened when you had the thought she snapped at me”