

04-Convert CSV to Txt

In this notebook, we will convert csv output of datasets (mini dev set, full dev set, test set) to txt format suitable for eval code

```
In [1]: import pandas as pd
```

```
In [2]: # list of files to convert
filenames = [
    "./data/output/mini_dev_set/elife_groq_mistral_summary.csv",
    "./data/output/mini_dev_set/plos_groq_mistral_summary.csv",
    "./data/output/full_dev_set/elife_groq_mistral_summary.csv",
    "./data/output/full_dev_set/plos_groq_mistral_summary.csv",
    "./data/output/test_set/elife_groq_mistral_summary.csv",
    "./data/output/test_set/plos_groq_mistral_summary.csv"
]

output_filenames = [
    "./data/output/mini_dev_set/elife.txt",
    "./data/output/mini_dev_set/plos.txt",
    "./data/output/full_dev_set/elife.txt",
    "./data/output/full_dev_set/plos.txt",
    "./data/output/test_set/elife.txt",
    "./data/output/test_set/plos.txt"
]
```

```
In [3]: for fname, output_fname in zip(filenames, output_filenames):
    print("Processing file =", fname)
    df = pd.read_csv(fname)
    print("Len =", len(df))

    print("Writing to file =", output_fname)
    # does some simple data cleaning
    df_txt = df["groq_mistral_summary"].apply(lambda text: text.replace("\n", ""))
    df_txt.to_csv(output_fname,
                  header=False,
                  index=False,
                  sep="\n"
    )
    print("Completed")
    print("-----")

print("=====")
print("All completed")
```

```
Processing file = ./data/output/mini_dev_set/elife_groq_mistral_summary.csv
Len = 24
Writing to file = ./data/output/mini_dev_set/elife.txt
Completed
-----
Processing file = ./data/output/mini_dev_set/plos_groq_mistral_summary.csv
Len = 138
Writing to file = ./data/output/mini_dev_set/plos.txt
Completed
-----
Processing file = ./data/output/full_dev_set/elife_groq_mistral_summary.csv
Len = 241
Writing to file = ./data/output/full_dev_set/elife.txt
Completed
-----
Processing file = ./data/output/full_dev_set/plos_groq_mistral_summary.csv
Len = 1376
Writing to file = ./data/output/full_dev_set/plos.txt
Completed
-----
Processing file = ./data/output/test_set/elife_groq_mistral_summary.csv
Len = 142
Writing to file = ./data/output/test_set/elife.txt
Completed
-----
Processing file = ./data/output/test_set/plos_groq_mistral_summary.csv
Len = 142
Writing to file = ./data/output/test_set/plos.txt
Completed
-----
=====
All completed
```

In []: