Cleaner.py Documentation

The cleaner.py script processes a CSV file with raw channel messages and performs several data cleaning and transformation operations on it. The resulting cleaned and transformed data is then saved to a new CSV file.

The script uses the pandas library to read the raw CSV file into a DataFrame. It then applies the following cleaning and transformation operations:

1. Extract the channel ID from the peer\_id column.
2. Extract information from the from\_id and media columns into separate columns.
3. Extract information from the entities, reply\_to, and replies columns into separate columns.
4. Strip the sizes, replies, entities, and reply\_to columns from the DataFrame.
5. Extract information from the sizes column into separate columns.
6. Finally, the cleaned and transformed DataFrame is saved to a new CSV file.

It is important to note that the script assumes that the input CSV file has a specific structure and format. If the input file has a different structure, the script may not work correctly.

Note file encoding must be UTF-8