ugen3.py

Client: PANTHEON.tech
Author: Marek Paulik
Date: 2024-06-16 23:14

Command line tool for generating usernames. Input data are provided from one or more formatted input files. Processed data are formatted and stored in specified output file.

Disclaimer: For using test.py, please set *usage* variable to "test" in *FormatCli* class (ugen3.py: line 18) for correct formatting of CLI output in this report file. For use in production, set the *usecase* variable to "production". If set to "test" in production, script will work as intended, only CLI output will be less fancy. :)

Usage: Input files must be named *input_file1.txt*, *input_file2.txt*, ... *input_fileN.txt* encoded in utf-8. Format of the data is *ID, Forename, Middle Name (optional), Surname, Department* separated by colon.

Example input data line:

```
1234:Jozef:Miloslav:Hurban:Legal
```

Script is validating input data. If *ID* is not an integer, if *Forename* or *Surname* is missing, it will skip the line and report it after running.
If *Department* is missing, script will process the line, but also report it after running.

To run the script you have to provide output file using *-o* or *- -output* flag and input file(s) using positional arguments.

Example CLI command:

```
python3 ugen3.py -o output_file.txt input_file1.txt input_file2.txt input_file3.txt
```

Running script with *-h* or *- -help* flag will display help.
If required arguments are missing or incorrect (wrong filepath, missing input file...) it will display error and help and exit.

Test case 1

Running script without arguments.

CLI command:

```
python3 ugen3.py
```

CLI Output:
```
!!! ERROR !!!
the following arguments are required: FILEPATH
!!! ERROR !!!


usage: ugen.py [-h] [-o PATH] FILEPATH [FILEPATH ...]

Generate list of usernames in single output file based on user info from
multiple input files.

positional arguments:
  FILEPATH              filename(s) of user info file(s)

options:
  -h, --help            show this help message and exit
  -o PATH, --output PATH
                        Specify the name of the output file.
```

Test case 2

Running script with valid arguments and correctly formatted input files.

Input file content:
```
1234:Jozef:Miloslav:Hurban:Legal
4567:Milan:Rastislav:Stefanik:Defence
4563:Jozef::Murgas:Development
```

CLI command:
```
python3 ugen3.py -o test/test_output.txt test/test_input.txt
```

CLI Output:
Output file content:
```
1234:jmhurban:Jozef:Miloslav:Hurban:Legal
4567:mrstefan:Milan:Rastislav:Stefanik:Defence
4563:jmurgas:Jozef::Murgas:Development
```

Test case 3

Script is handling length of usernames (8 chars) and duplicate usernames (adding number to username).

Input file content:

```
4567:Milan:Rastislav:Stefanik:Defence
1111:Pista::Hufnagel:Sales
4563:Pista::Hufnagel:Sales
```

CLI command:
```
python3 ugen3.py -o test/test_output.txt test/test_input.txt
```

CLI Output:
Output file content:
```
4567:mrstefan:Milan:Rastislav:Stefanik:Defence
1111:phufnage:Pista::Hufnagel:Sales
4563:phufnage1:Pista::Hufnagel:Sales
```

Test case 4

Providing ill-formatted data will result in only partial processing.
Possible format errors:
(Lines in input file are corresponding to list below.)
- wrong number of values in line
- using different separator
- ID is not a number
- missing Forename
- missing Surname
- missing Department
(Last line is correctly formatted)

Input file content:
```
1234:Miloslav:Hurban:Legal
7890,Milan,Rastislav,Stefanik,Defence
abcd:Milan:Rastislav:Stefanik:Defence
4563:::Murgas:Development
1111:Pista:::Sales
4563:Pista::Hufnagel:
7890:Milan:Rastislav:Stefanik:Defence
```

CLI command:
```
python3 ugen3.py -o test/test_output.txt test/test_input.txt
```

CLI Output:
```
Traceback (most recent call last):
  File "/mnt/data/_nextcloud/work/code/pantheon/240614_homework/ugen3.py",
line 57, in <module>
    input_df["Forename"].apply(lambda s: s[0].lower())
  File
"/mnt/data/_nextcloud/work/code/pantheon/240614_homework/.venv/lib/python3
.10/site-packages/pandas/core/series.py", line 4924, in apply
```

```
    ).apply()
  File
"/mnt/data/_nextcloud/work/code/pantheon/240614_homework/.venv/lib/python3
.10/site-packages/pandas/core/apply.py", line 1427, in apply
    return self.apply_standard()
  File
"/mnt/data/_nextcloud/work/code/pantheon/240614_homework/.venv/lib/python3
.10/site-packages/pandas/core/apply.py", line 1507, in apply_standard
    mapped = obj._map_values(
  File
"/mnt/data/_nextcloud/work/code/pantheon/240614_homework/.venv/lib/python3
.10/site-packages/pandas/core/base.py", line 921, in _map_values
    return algorithms.map_array(arr, mapper, na_action=na_action,
convert=convert)
  File
"/mnt/data/_nextcloud/work/code/pantheon/240614_homework/.venv/lib/python3
.10/site-packages/pandas/core/algorithms.py", line 1743, in map_array
    return lib.map_infer(values, mapper, convert=convert)
  File "lib.pyx", line 2972, in pandas._libs.lib.map_infer
  File "/mnt/data/_nextcloud/work/code/pantheon/240614_homework/ugen3.py",
line 57, in <lambda>
    input_df["Forename"].apply(lambda s: s[0].lower())
TypeError: 'NoneType' object is not subscriptable
```

Output file content:

Test case 5

Testing possible format errors one by one.

1. Wrong number of values in line

Input file content:
1234:Miloslav:Hurban:Legal

CLI command:
python3 ugen3.py -o test/test_output.txt test/test_input.txt

CLI Output:
```
Traceback (most recent call last):
  File
"/mnt/data/_nextcloud/work/code/pantheon/240614_homework/.venv/lib/python3
.10/site-packages/pandas/core/internals/construction.py", line 939, in
_finalize_columns_and_data
    columns = _validate_or_indexify_columns(contents, columns)
  File
```

```
"/mnt/data/_nextcloud/work/code/pantheon/240614_homework/.venv/lib/python3
.10/site-packages/pandas/core/internals/construction.py", line 986, in
_validate_or_indexify_columns
    raise AssertionError(
AssertionError: 5 columns passed, passed data had 4 columns

The above exception was the direct cause of the following exception:

Traceback (most recent call last):
  File "/mnt/data/_nextcloud/work/code/pantheon/240614_homework/ugen3.py",
line 51, in <module>
    input_df = pd.DataFrame(input_data, columns=column_names)
  File
"/mnt/data/_nextcloud/work/code/pantheon/240614_homework/.venv/lib/python3
.10/site-packages/pandas/core/frame.py", line 851, in __init__
    arrays, columns, index = nested_data_to_arrays(
  File
"/mnt/data/_nextcloud/work/code/pantheon/240614_homework/.venv/lib/python3
.10/site-packages/pandas/core/internals/construction.py", line 520, in
nested_data_to_arrays
    arrays, columns = to_arrays(data, columns, dtype=dtype)
  File
"/mnt/data/_nextcloud/work/code/pantheon/240614_homework/.venv/lib/python3
.10/site-packages/pandas/core/internals/construction.py", line 845, in
to_arrays
    content, columns = _finalize_columns_and_data(arr, columns, dtype)
  File
"/mnt/data/_nextcloud/work/code/pantheon/240614_homework/.venv/lib/python3
.10/site-packages/pandas/core/internals/construction.py", line 942, in
_finalize_columns_and_data
    raise ValueError(err) from err
ValueError: 5 columns passed, passed data had 4 columns
```

Output file content:

2. Using different separator

Input file content:
```
7890,Milan,Rastislav,Stefanik,Defence
```

CLI command:
```
python3 ugen3.py -o test/test_output.txt test/test_input.txt
```

CLI Output:
```
Traceback (most recent call last):
  File
```

```
"/mnt/data/_nextcloud/work/code/pantheon/240614_homework/.venv/lib/python3
.10/site-packages/pandas/core/internals/construction.py", line 939, in
_finalize_columns_and_data
    columns = _validate_or_indexify_columns(contents, columns)
  File
"/mnt/data/_nextcloud/work/code/pantheon/240614_homework/.venv/lib/python3
.10/site-packages/pandas/core/internals/construction.py", line 986, in
_validate_or_indexify_columns
    raise AssertionError(
AssertionError: 5 columns passed, passed data had 1 columns

The above exception was the direct cause of the following exception:

Traceback (most recent call last):
  File "/mnt/data/_nextcloud/work/code/pantheon/240614_homework/ugen3.py",
line 51, in <module>
    input_df = pd.DataFrame(input_data, columns=column_names)
  File
"/mnt/data/_nextcloud/work/code/pantheon/240614_homework/.venv/lib/python3
.10/site-packages/pandas/core/frame.py", line 851, in __init__
    arrays, columns, index = nested_data_to_arrays(
  File
"/mnt/data/_nextcloud/work/code/pantheon/240614_homework/.venv/lib/python3
.10/site-packages/pandas/core/internals/construction.py", line 520, in
nested_data_to_arrays
    arrays, columns = to_arrays(data, columns, dtype=dtype)
  File
"/mnt/data/_nextcloud/work/code/pantheon/240614_homework/.venv/lib/python3
.10/site-packages/pandas/core/internals/construction.py", line 845, in
to_arrays
    content, columns = _finalize_columns_and_data(arr, columns, dtype)
  File
"/mnt/data/_nextcloud/work/code/pantheon/240614_homework/.venv/lib/python3
.10/site-packages/pandas/core/internals/construction.py", line 942, in
_finalize_columns_and_data
    raise ValueError(err) from err
ValueError: 5 columns passed, passed data had 1 columns
```

Output file content:

3. ID is not a number
Input file content:
abcd:Milan:Rastislav:Stefanik:Defence

CLI command:
python3 ugen3.py -o test/test_output.txt test/test_input.txt

Output file content:
```
abcd:mrstefan:Milan:Rastislav:Stefanik:Defence
```

4. Missing Forename
Input file content:
```
4563:::Murgas:Development
```

CLI command:
```
python3 ugen3.py -o test/test_output.txt test/test_input.txt
```

CLI Output:
```
Traceback (most recent call last):
  File "/mnt/data/_nextcloud/work/code/pantheon/240614_homework/ugen3.py",
line 57, in <module>
    input_df["Forename"].apply(lambda s: s[0].lower())
  File
"/mnt/data/_nextcloud/work/code/pantheon/240614_homework/.venv/lib/python3
.10/site-packages/pandas/core/series.py", line 4924, in apply
    ).apply()
  File
"/mnt/data/_nextcloud/work/code/pantheon/240614_homework/.venv/lib/python3
.10/site-packages/pandas/core/apply.py", line 1427, in apply
    return self.apply_standard()
  File
"/mnt/data/_nextcloud/work/code/pantheon/240614_homework/.venv/lib/python3
.10/site-packages/pandas/core/apply.py", line 1507, in apply_standard
    mapped = obj._map_values(
  File
"/mnt/data/_nextcloud/work/code/pantheon/240614_homework/.venv/lib/python3
.10/site-packages/pandas/core/base.py", line 921, in _map_values
    return algorithms.map_array(arr, mapper, na_action=na_action,
convert=convert)
  File
"/mnt/data/_nextcloud/work/code/pantheon/240614_homework/.venv/lib/python3
.10/site-packages/pandas/core/algorithms.py", line 1743, in map_array
    return lib.map_infer(values, mapper, convert=convert)
  File "lib.pyx", line 2972, in pandas._libs.lib.map_infer
  File "/mnt/data/_nextcloud/work/code/pantheon/240614_homework/ugen3.py",
line 57, in <lambda>
    input_df["Forename"].apply(lambda s: s[0].lower())
IndexError: string index out of range
```

Output file content:

5. Missing Surname

Input file content:

```
1111:Pista:::Sales
```

CLI command:

```
python3 ugen3.py -o test/test_output.txt test/test_input.txt
```

CLI Output:
Output file content:

```
1111:p:Pista:::Sales
```

6. Missing Deaprtment

Input file content:

```
4563:Pista::Hufnagel:
```

CLI command:

```
python3 ugen3.py -o test/test_output.txt test/test_input.txt
```

CLI Output:
Output file content:

```
4563:phufnage:Pista::Hufnagel:
```

Test case 6

If incorrect data is provided, nothing is processed.

Input file content:

```
Lorem ipsum dolor sit amet, consectetur adipiscing elit. Suspendisse
auctor enim ac velit ultrices euismod. Praesent felis nisi, tempor ac
posuere eu, interdum a magna. Quisque gravida at felis ut viverra.
Vestibulum eget dignissim lorem,
vel viverra nisi. Proin est risus, blandit at arcu vel, rutrum
pellentesque nulla. Aenean accumsan elit libero. Donec ut auctor nisl.
Aenean faucibus nisi sapien, in placerat nulla cursus eget. Nullam euismod
massa vitae dapibus tincidunt.
Sed porttitor
fermentum
blandit.
Praesent aliquet mi vulputate augue pharetra commodo. Maecenas lacus
tortor, tristique sit amet pretium in, fermentum quis augue. Proin
efficitur venenatis ligula ac mattis.
```

CLI command:

```
python3 ugen3.py -o test/test_output.txt test/test_input.txt
```

```
Traceback (most recent call last):
  File
"/mnt/data/_nextcloud/work/code/pantheon/240614_homework/.venv/lib/python3
.10/site-packages/pandas/core/internals/construction.py", line 939, in
_finalize_columns_and_data
    columns = _validate_or_indexify_columns(contents, columns)
  File
"/mnt/data/_nextcloud/work/code/pantheon/240614_homework/.venv/lib/python3
.10/site-packages/pandas/core/internals/construction.py", line 986, in
_validate_or_indexify_columns
    raise AssertionError(
AssertionError: 5 columns passed, passed data had 1 columns

The above exception was the direct cause of the following exception:

Traceback (most recent call last):
  File "/mnt/data/_nextcloud/work/code/pantheon/240614_homework/ugen3.py",
line 51, in <module>
    input_df = pd.DataFrame(input_data, columns=column_names)
  File
"/mnt/data/_nextcloud/work/code/pantheon/240614_homework/.venv/lib/python3
.10/site-packages/pandas/core/frame.py", line 851, in __init__
    arrays, columns, index = nested_data_to_arrays(
  File
"/mnt/data/_nextcloud/work/code/pantheon/240614_homework/.venv/lib/python3
.10/site-packages/pandas/core/internals/construction.py", line 520, in
nested_data_to_arrays
    arrays, columns = to_arrays(data, columns, dtype=dtype)
  File
"/mnt/data/_nextcloud/work/code/pantheon/240614_homework/.venv/lib/python3
.10/site-packages/pandas/core/internals/construction.py", line 845, in
to_arrays
    content, columns = _finalize_columns_and_data(arr, columns, dtype)
  File
"/mnt/data/_nextcloud/work/code/pantheon/240614_homework/.venv/lib/python3
.10/site-packages/pandas/core/internals/construction.py", line 942, in
_finalize_columns_and_data
    raise ValueError(err) from err
ValueError: 5 columns passed, passed data had 1 columns
```