

20_drop_young_writers

March 21, 2017

02/26/17 - smiel ### Very young writers are also sort of language learners, so we want to uninclude them from this analysis

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

path = '/research/ella/rivendell'

In [2]: # if original versions exist, use those
orig = '.orig_20'

# essays
essays = os.path.join(path, 'all_essays.csv')
if os.path.exists(essays + orig):
    df = pd.read_csv(essays + orig, encoding='utf8')
else:
    df = pd.read_csv(essays, encoding='utf8')

print('{} essays'.format(len(df)))

# make a copy
df.to_csv(essays + orig, encoding='utf8', index=False)

# drop young writers
df.drop(df.index[df.age < 13], inplace=True)

# save
df.to_csv(essays, encoding='utf8', index=False)

print('{} essays'.format(len(df)))

# sentences
sentences = os.path.join(path, 'all_sentences.csv')
if os.path.exists(sentences + orig):
    df = pd.read_csv(sentences + orig, encoding='utf8')
else:
    df = pd.read_csv(sentences, encoding='utf8')
```

```

print('{} sentences'.format(len(df)))

# make a copy
df.to_csv(sentences + orig, encoding='utf8', index=False)

# drop young writers
df.drop(df.index[df.age < 13], inplace=True)

# save
df.to_csv(sentences, encoding='utf8', index=False)

print('{} sentences'.format(len(df)))

/home/smiel/.venvs/rivendell/local/lib/python2.7/site-packages/IPython/core/interactiveshell.py:
    interactivity=interactivity, compiler=compiler, result=result)

141520 essays
111755 essays

/home/smiel/.venvs/rivendell/local/lib/python2.7/site-packages/IPython/core/interactiveshell.py:
    interactivity=interactivity, compiler=compiler, result=result)

2382191 sentences
1870084 sentences

```