

Upgoer

This exercise is based on the XKCD comic/book called “Upgoer” (<https://xkcd.com/1133/>) that attempts to explain concepts using only the 1000-most common English words. Write a Python program called `upgoer.py` that reads a `-w|--wordlist` file of those 1000 common words and takes a single positional argument that should be a filename. Read all the words in that file and print each word *to STDERR* in the file that is not found in the wordlist file. Prefix each word with the number of the word (e.g., the first word has “1:” but use `{:3}: {}` to print the number and the word). You do not need to create a unique list of words to change; that is, if you see “heffalump” three times in a document, print and count it three times in the output. Finally print the total number of words that need to be changed.

You should consider the words in a case-insensitive fashion (lowercase them all) and remove any characters that are not numbers or letters or the apostrophe. This would be useful to you:

```
import re
re.sub("[^a-zA-Z0-9']", '', word)
```

Expected Output

```
$ ./upgoer.py
usage: upgoer.py [-h] [-w str] FILE
upgoer.py: error: the following arguments are required: FILE
$ ./upgoer.py -h
usage: upgoer.py [-h] [-w str] FILE
```

Find words in text not in 1000 most-common

```
positional arguments:
  FILE                File to check
```

```
optional arguments:
  -h, --help            show this help message and exit
  -w str, --wordlist str
                        Common word file (default: 1000.txt)
```

```
$ ./upgoer.py lkdf
"lkdf" is not a file
$ ./upgoer.py hello.txt
1: hello
1 word to change.
$ ./upgoer.py hello.txt 2>err
1 word to change.
```

```
$ cat err
1: hello
$ ./upgoer.py nobody.txt
1: i'm
2: nobody
3: nobody
4: theres
5: dont
6: theyd
7: advertise
8: dreary
9: somebody
10: public
11: frog
12: one's
13: livelong
14: june
15: admiring
16: bog
16 words to change.
```

Test Suite

A passing test suite looks like this:

```
$ make test
pytest -v test.py
===== test session starts =====
platform darwin -- Python 3.6.8, pytest-4.2.0, py-1.7.0, pluggy-0.8.1 -- /anaconda3/bin/python
cachedir: .pytest_cache
rootdir: /Users/kyclark/work/worked_examples/2019_spring_finals/upgoer, inifile:
plugins: remotedata-0.3.1, openfiles-0.3.2, doctestplus-0.2.0, arraydiff-0.3
collected 6 items

test.py::test_usage PASSED [ 16%]
test.py::test_bad_wordlist PASSED [ 33%]
test.py::test_bad_input PASSED [ 50%]
test.py::test_hello PASSED [ 66%]
test.py::test_nobody PASSED [ 83%]
test.py::test_gettysburg PASSED [100%]

===== 6 passed in 0.64 seconds =====
```