C_HW8 (common words)

Ken Youens-Clark

Abstract

Find the words in common between two files. It is suggested to use hashes, but Sets and Bags would also useful for the more advanced student.

Protocol

Step 1.

Checkout or update (git pull) https://github.com/kyclark/metagenomics-book

$ ssh hpc

$ (cd metagenomics-book && git pull)

$ cp -r metagenomics-book/problems/common-words /abe487/problems

$ cd !$

$ git add common-words

$ git commit -m 'adding common words' common-words

$ git push

Step 2.

Write the code for 'common.pl6' that finds the words that are in common between two files. You are provide three pairs of files with the following known common words:

const.txt (US Constitution) dec.txt (Declaration of Independence) = 862

foo bar = 0

text1.txt text2.txt = 8

You will need to lowercase words to compare them, and be sure to remove any punctuation. You will find this useful:

```
for 'const.txt'.IO.lines -> $line {
    dd $line.subst(/<-[\w\s]>/, '', :g).words;
}
```

That regular expression says to create a character class of word-characters (\w) and space-like characters (\s) and negate it ("-"). Use that as the matcher to substitute ("subst") those matching with the empty string ("").

Step 3.

A passing test suite looks like this:

```
$ make test
./test.pl6
ok 1 - No args gives usage
ok 2 - dec/const = 262
ok 3 - foo/bar = 0
ok 4 - text1/text2 = 8
```