



THORGATE

Python CodeClub

30.09.2015

TTÜ innovatsiooni-ja ettevõtluskeskus
MEKTORY
MODERN ESTONIAN KNOWLEDGE TRANSFER ORGANIZATION FOR YOU





Python CodeClub
bit.ly/pythoncodeclub

Questions from last time

CodeClub mentors

- Help others and present your own challenges
- Special T-shirt to stand out from the others (soon :)
- Let us know and put your name down
- Try to come 30 minutes before next CodeClub starts

Today's challenge - files and file system

Create a program that accepts file names as arguments,
parses the files and
creates CSV files with *interesting* facts about the files
(try to be creative and write your own solution)

```
python parse.py file1.txt file2.txt ...
```


Input files

You can use Samuel L. Ipsum to create the input files

<http://slipsum.com/>

Things to consider

- File endings
- Different cases of letters
- Ignoring punctuation
- Unicode characters
- Ignoring binary files (that don't have text)

Output 1 - CSV file

Output the number of different words in each file

Example output1.csv

file1.txt,123

file2.txt,1830

...

Output 2 - CSV file

Output the number of most common words that occur in all files together in descending order

Example output2.csv

and,28948

to,12346

CodeClub,3

...

Output 3 - CSV file

Try to implement your own spell checker and print out words that are most probably spelled wrong based on single-letter mutation and swapping events. Output the word together with the file name and line number where they occur.

Example output3.csv

```
taemwork,file1.txt,23  
taemwork,file2.txt,1465  
winz,file2.txt,3  
championchips,file3.txt,1
```




bit.ly/pythoncodeclub