

SPAM filter

Jan Sadílek a Kateřina Rejchrtová

Úvod





Filter.py

basefilter.py

corpus.py

spam_words.py

(utils.py)

Postup



Generování
mailů

corpus.py



Rozdělení
obsahu do
slov

basefilter.py



SPAM slova

spam_words.py



Počítání
shodných
slov

filter.py



Vyhodnocení

filter.py

Vyhodnocení

```
def decide(self, fname, body):  
    self.count_to_dic(body.lower().split())  
    self.spam_word_count = sum(self.spam_word_dict.values())  
    if self.spam_word_count > 0:  
        return "SPAM"  
    else:  
        return "OK"
```

spam_words.py

```
SPAM_WORDS = ["own home based business", "perfect",  
"paid", "fortune", "win", "font", "spam", "shape up for",  
"viagra", "weight-loss", "adult site", "special promotion",  
"promise", "expect to earn", "congratulations", "this is not a  
joke", "credit card information"]
```



Děkujeme za pozornost!

Zdroje

- <https://cw.fel.cvut.cz/wiki/courses/b4b33rph/cviceni/spam/start>
- testíky z hodin Řešení problémů a hry
- <https://en.wikipedia.org/wiki/Spamming>