

作业

单词统计增加忽略单词

对sample文件进行不区分大小写的单词统计。

要求用户可以排除一些单词的统计，例如a、the、of等不应该出现在具有实际意义的统计中，应当忽略。

要求，全部代码使用函数封装、调用完成。

```
def makekey2(line:str, chars=set('"'!'"()[\]{},*-/.\&: \r\n')):
    """处理行，分离出单词

    :param line: 行字符串
    :param chars: 分隔符，包括空白字符
    :return: 生成器，分离出的单词
    """
    start = 0
    for i, c in enumerate(line):
        if c in chars: # 如果是分隔符，才有可能切片
            if start == i: # 如果紧挨着还是特殊字符，start一定等于i
                start += 1
                continue
            yield line[start:i]
            start = i + 1 # 加1是跳过这个不需要的特殊字符c
    else:
        if start < len(line): # 小于，说明还有有效的字符，而且一直到末尾
            yield line[start:]

#print(list(makekey2(':::abc,.')))
#print(list(makekey2('-----'))))

def wordcount(filename, encoding='utf-8', ignore=set()):
    d = {}
    with open(filename, encoding=encoding) as f:
        for line in f:
            for word in map(str.lower, makekey2(line)):
                if word not in ignore:
                    d[word] = d.get(word, 0) + 1
    print(d)
    return d

def top(d:dict, n=10):
    yield from filter(lambda x: x[0]<=n, enumerate(sorted(d.items(), key=lambda item:item[1],
reverse=True), 1))

# 单词统计前几名
t = top(wordcount('sample.txt', ignore={'the', 'is', 'a'}))
print(list(t))
```

转换为json文件

有一个配置文件test.ini内容如下，将其转换成json格式文件

```
[DEFAULT]
a = test

[mysql]
default-character-set=utf8
a = 1000

[mysqld]
datadir = /dbserver/data
port = 33060
character-set-server=utf8
sql_mode=NO_ENGINE_SUBSTITUTION,STRICT_TRANS_TABLES
```

遍历ini文件的字典即可

```
from configparser import ConfigParser
import json

filename = 'test.ini'
jsonname = 'test.json'

cfg = ConfigParser()
cfg.read(filename)

dest = {}

for sect in cfg.sections():
    print(sect, cfg.items(sect))
    dest[sect] = dict(cfg.items(sect))

json.dump(dest, open(jsonname, 'w'))

# 可以使用下面的命令格式化json      python -m json.tool test.json
```