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

之前代码中，切分单词的太繁琐，因为makekey1函数已经可以直接把一行数据切成一个个单词了，所以对上面的代码重新封装。

```
def makekey2(line:str, chars=set("""!'"#./\()[],*- \r\n""")):
    start = 0

    for i, c in enumerate(line):
        if c in chars:
            if start == i: # 如果紧挨着还是特殊字符，start一定等于i
                start += 1 # 加1并continue
                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:]

def wordcount(filename, encoding='utf8', 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
    return d

def top(d:dict, n=10):
    for i, (k,v) in enumerate(sorted(d.items(), key=lambda item: item[1], reverse=T
rue)):
        if i > n:
```

```
            break
        print(k,v)

# 单词统计前几名
top(wordcount('sample', ignore={'the', 'a'}))
```

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'))

# $ python -m json.tool test.json
```