

csv文件

csv文件简介

参看 RFC 4180 <http://www.ietf.org/rfc/rfc4180.txt>

CSV即逗号分隔值Comma-Separated Values

- 它是一个被行分隔符、列分隔符划分成行和列的文本文件
- CSV 不指定字符编码
- 每一行称为一条记录record
- 表头可选，和字段列对齐就行了

分隔符

- 行分隔符为\r\n，最后一行可以没有换行符
- 列分隔符常为逗号或者制表符

字段

- 字段可以使用双引号括起来，也可以不使用
- 如果字段中出现了双引号、逗号、换行符必须使用双引号括起来
- 如果字段的值是双引号，使用两个双引号表示一个转义

csv 模块

`csv.reader(csvfile, dialect='excel', **fmtparams)` 返回reader对象，是一个行迭代器。

默认使用excel方言，如下：

- delimiter 列分隔符，逗号
- lineterminator 行分隔符 \r\n
- quotechar 字段的引用符号，缺省为 " 双引号
- 双引号的处理
 - doublequote 双引号的处理，默认为True。如果碰到数据中有双引号，而quotechar也是双引号，True则使用2个双引号表示，False表示使用转义字符将作为双引号的前缀
 - escapechar 一个转义字符，默认为None
 - writer = csv.writer(f, doublequote=False, escapechar='@') 遇到双引号，则必须提供转义字符
- quoting 指定双引号的规则
 - QUOTE_ALL 所有字段
 - QUOTE_MINIMAL 特殊字符字段，Excel方言使用该规则
 - QUOTE_NONNUMERIC 非数字字段
 - QUOTE_NONE 都不使用引号。

`csv.writer(csvfile, dialect='excel', **fmtparams)` 返回DictWriter的实例。

主要方法有单行writerow、多行writerows。

```
1 import csv
2
3 rows = [
```

```

4      ('id', 'name', 'age', 'comment'),
5      [1, 'tom', 20, 'tom'],
6      (2, 'jerry', 18, 'jerry'),
7      (3, 'justin', 22, 'just\t"in'),
8      "abcdefgh",
9      ((1,), 2, [3])
10 ]
11
12 # newline='', 表示写入时, 不要做\n的替换, 那么输出就是\r\n
13 with open('o:/test.csv', 'w+', encoding='utf-8', newline='') as f:
14     writer = csv.writer(f)
15     writer.writerow(rows[0])
16     writer.writerows(rows[1:])
17
18 # newline='', 表示读取时, 也不做\n的替换
19 with open('o:/test.csv', encoding='utf-8', newline='') as f:
20     reader = csv.reader(f) # 行迭代器
21     print(next(reader))
22     print(next(reader))
23     for line in reader:
24         print(line)

```

每一行写入一个可迭代对象, 可迭代的每一个元素, 将作为csv行中的每一列的元素。

windows下在会在每行末尾多出一个\r, 解决办法 `open('test.csv', 'w', newline='')`