

Q1：有其他的库能更快捷更高效访问所存的数据吗？

一般来说pandas封装的函数能帮助更快捷更高效的处理所存二维数据，若数据结构更为复杂，使用数据库和SQL语句更为高效、便捷。

Q2：为什么cat a.xlsx文件会乱码，是不是csv和xlsx的编码方式不同造成的？是否需要先转化编码格式？

cat指令只能有效读取文本文件，而xlsx是Microsoft自定义的一种格式，准备来说是文件编码方式不同，而不是我们通常说的字符编码格式(字符编码跟你的系统和软件编码设置有关)。

Q3：pandas读取文件时怎么使用正则表达式来指定分隔符？

pd.read_csv()等读取文件函数有两个参数sep和delimiter sep : str, default ',' 指定分隔符，可以为正则表达式，例如'\r\t'。

delimiter : str, default None 定界符，备选分隔符（如果指定该参数，则sep参数失效） 若需要对字符串做更复杂的操作，需要使用更基本的库。

关于分隔符处理

- csv module只能处理单个分隔符；
- 需处理多个分隔符，使用string.split
- 更复杂的操作使用正则表达式:re.split

Q4：data[0]['title']中的title与columns=['title']有哪些区别。[0]是指的一个表还是什么意思？

data是一个列表，data[0]是一个字典，columns指定构建DataFrame的列。

其实很好理解，因为之前有用list of dict创建DataFrame，现在只不过指定了列(columns)