

## 第20课三国演义之谁是话痨

作业辅导：每周6次

笔记分享

python课程=语法课程+面向对象课程（课少）+系统学习课（小组式）

周日考试，9点开始（不能迟到），先完先走

第二轮录屏开始

——摸底开始

新的课程在路上-----

### 课程计划

Python基础课程-->C语言应用课（嵌入式Arduino）-->C语言算法课 -->C++高级应用课 --->人工智能编程（或嵌入式应用）

附加：C语言高级同步课（大一下），Python面向对象同步课（大二上）

课程特点：实用性、动手能力强。起码保你能找到事做，优秀者保你进大厂（或保985、211）

第16周后停课复习

## 新知识点

学习操作文件指针

(1) `seek(offset[,whence])`:

(2) `offset`--偏移量，可以是负值，代表从后向前移动；

(3) `whence`--偏移相对位置，分别有：`os.SEEK_SET`（相对文件起始位置，也可用“0”表示）；

`os.SEEK_CUR`（相对文件当前位置，也可用“1”表示）；`os.SEEK_END`（相对文件结尾位置，也可用“2”表示）。

`seek(x, 0)`：表示指针从开头位置移动到x位置

`seek(x, 1)`：表示指针从当前位置向后移动x个位置

`seek(-x, 2)`：表示指针从文件结尾向前移动x个位置

例：`file.seek(-1,2)`，文件指针从文件末尾向前移动一个字符，配合`read`相关方法/函数可读取该字符。

学习读取文件的部分内容

`read(size)`

通过使用 `size` 参数，指定 `read()` 每次可读取的最大字符（或者字节）数

`read([size])` 方法从文件当前位置起读取size个字节

注意：`size`，包括换行符

认识文本文件：深刻认识文本文件的内部组成及二进制值编码（核心能力）

正则表达式：

1.[https://www.bilibili.com/video/BV1xs411x71b?from=search&seid=10820040791487531796&spm\\_id\\_from=333.337.0.0](https://www.bilibili.com/video/BV1xs411x71b?from=search&seid=10820040791487531796&spm_id_from=333.337.0.0)

2.<https://www.cnblogs.com/sirxy/p/11815169.html>

《三国演义》文本分析技术：

- 1) 归一化处理技术：每人选一个主角人物，对其归一化处理
- 2) 完整性处理：补全
- 3) 过滤技术：无对话无主角段落
- 4) 手工与编程相结合技术

## 作业任务（周六、下周二不上课）

### 文件操作任务

任务1：自定义一个含有中英文文字、空格回车、标点符号、数字外文等尽可能多文本内容的多段文字的文本文件，将其内容按字节"翻译"到excel表解读成字符和二进制（难度系数：底层操作，巨难）

**《三国演义》之谁是话痨任务：找出话痨(说话次数最多，说话字数最多) 强烈建议以小组为单位讨论**

任务2：找到《三国演义》之谁是话痨，并给出工作思路（md文档）、中间过程（多个程序及存盘数据文件）和最后统计结果数据（结果文档）（项目作业，工作量巨大）