

2021年春季学期 北京大学公共选修课
计算机科学与编程入门

第2讲 Python语言进阶

主讲教师：陆俊林 纪晓璐
2021年3月15日



主要内容



01. 列表



02. 字典



03. 库

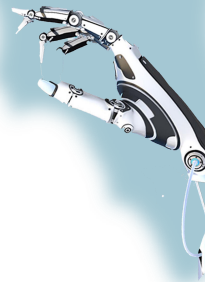


04. 读写文件

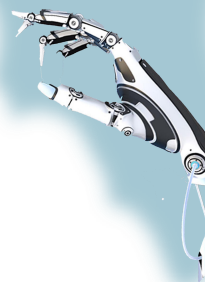


程序示例

- ④ 本讲的内容都在程序示例中
- ④ 认真阅读程序示例的代码和注释
- ④ 对照代码和运行结果，掌握具体功能
- ④ 尝试自己编写类似功能的代码



作业（内容）

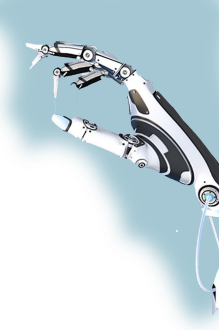


- ④ 选择一个长篇文本，作为分析对象
 - 十万字以上为宜，可以是小说、剧本、新闻纪实等，科研著作也可以
 - 分析对象应是自己熟悉的领域，要分析选择该分析对象的原因

- ④ 参照课上讲授的方法，编写程序，进行词频统计
 - 统计文本中的某一类词，例如人名、地名等
 - 对于有可能产生误差之处，要有细致的分析和合理的应对
 - 比如：“诸葛亮”和“孔明”是否算同一个人物，“孔明曰”的数量要不要计入

- ④ 对统计结果进行分析

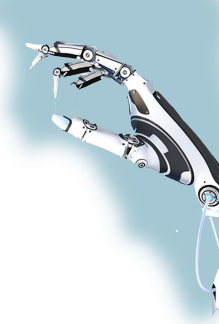
作业（评分）



🎯 评分要点

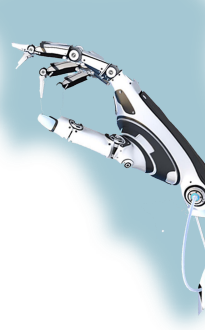
- 代码功能正确，结构清晰，语句运用合理
- 变量命名规范、简明、易懂
- 重要功能代码都有注释，内容丰富、要点明确、易于阅读
- 用户互动（输入输出）在准确的前提下，操作简洁，容错性高
- 说明文档内容翔实，版式美观

作业（提交格式）



- ④ 参考课程的程序示例，具体要求如下：
 - Python程序：1个py文件（不要分成多个.py文件）
 - 说明文档：1个docx文件，可以再附1个内容相同的pdf文件
 - 用于分析的长篇文本放在data目录下
 - 统计结果文件放在output目录下，文件格式为csv
 - 整个目录以自己学号命名，将目录打包为zip文件提交

作业（提交时间等）



🎯 提交时间

- 3月22日中午12:00之前提交
- 如确有困难，可最多延期48小时，但迟交会影响成绩
- 延期超过48小时，视为未交，本次作业没有成绩

🎯 其它

- 具体提交方式稍后通知
- 本次作业的结果还将作为后续可视化设计的基础
- 作业完成的出色，除了获得好成绩，还会有意外奖励惊喜

🎯 **最重要的：不要抄袭，包括别人的和网上的**

关于下一讲

- ④ 下一讲的主题是：可视化设计
- ④ 需要用到的Python第三方库
 - matplotlib (Anaconda已安装该库)
 - pyecharts



本讲到此结束，谢谢！

— 计算机科学与编程入门 —

