

实验 06 数据模型与类

实验内容: 练习并利用 Python 数据模型的 API 建立一个数据接口类, 将目标数据集封装成 Python 中的序列型数据; 并在主函数中实现类的实例化使用。

实验过程(自选 1 项*开头的题目, 必选):

*) 分析两本类型不同的现代小说的词性分布, 如武侠或侦探小说等. 用一个类读入正本小说, 用自然语言处理工具, 初始化过程分析内容, 分解词语并获取词性(如动词, 形容词等). 类对象取索引返回词和词性两项主要信息. 在调用类对象的函数中, 实现词性的统计, 用饼状图可视化个主要词性频率. 对比两本小说的饼状图.

*) 分析过去一段时间的手机照片拍照时间的统计数据或 GPS 信息. 用一个类来描述一个含有照片的文件目录; 取索引返回拍照时间或经纬度. 在主函数中遍历照片信息, 用两个图分别画出每天的拍照次数和一天各时间段的拍照频率; 或用在地图上用圆点标出全部照片的拍照地点. (地图绘制可留意 folium, basemap, bokeh 等包).

(选做)进一步, 依照拍照时间先后顺序, 在地图上画出带箭头的时空轨迹.

*) 自主设计实验. 关键实现利用 API 实现一个类, 用来做数据接口; 从主函数中调用类对象时, 将它看成一个序列型数据.

延伸实验(选做):

*) 可视化图像的小波通道(可能用到 pywt 模块).

*) 高帧率视频的分解与降速(可能用到 cv2 模块).

*) 音频高音消除, 环境噪声的降噪.

*) 建立一个小型的数据库, 存储文件名, 和对应校验和/哈希码(用到 hashlib 模块). 实现相关的文件校对函数.

*) 语音/视频口头禅检测与计数器. 检测类似'OK', '对吧?', '对不对?', '行不行'等. 大致流程: [语音/视频音轨提取]->[音频转文字(语音识别)]->[文字分析统计]. 注意: 难.