

任务板 - 进阶 - 使用python随机抽取文本内容构造公章图片

author: 杨宁 last modified: 2020-07-21

1. 概述

1. 任务内容

- 从文本文件中逐行读取内容，剔除空格、换行及不可见字符，最后合并所有文本内容
- 按「停用词表」剔除停用字词，如“逗号”、“的”等
- 随机抽取指定数量范围内的文本

2. 任务目的

随机生成公章图片用于训练

3. 任务等级：进阶

4. 任务时间：4日内

2. 环境要求

python 3.6

引用库：Pillow、numpy、opencv、math、random

3. 技术要点

1. 文本文件读取时，`utf-8`、`GB2312` 等字符编码问题的处理；
2. 中文字体库载入；
3. 不可见字符，如`'\u3000'`、`'\xa0'`等；
4. 标点符号等停用词处理；
5. 图像旋转及抗锯齿；
6. Python 自带 `random` 库中，`randint`，`samples` 等方法；
7. 自动标注坐标点，以及坐标旋转计算

4. 可演示内容



(待补充)

5. 常见问题处理

(待补充)

6. 遗留问题

- 加粗字体、细长字体等，原字体库不提供的类型处理