

# 大作业一: n-gram语言模型平滑算法

---

**截止时间** 2021 10月 17, 23:59 之前      **得分** 15      **提交** 一份上传文件  
**可用** 2021 9月 27 22:00 至 2021 10月 17 23:59 20 天

---

此作业锁定于 2021 10月 17 23:59。

## 数据集

- 下载链接: [hw1\\_dataset.zip](#)
- 相关介绍:
  - 格式: 纯文本
  - 规模: 共1700w个词汇, 按照8 : 1 : 1进行train : dev : test划分

## 任务描述

- 设计并实现一种discounting算法, 在制定的训练语料上构建语言模型, 并在后续给出的测试集上测试PPL
- 评分标准: 报告 + PPL测试结果

## 注意事项

- 鼓励使用各种开源工具, 但要在报告中注明自己利用了哪些工具, 自己实现的部分是哪些
- 鼓励与其他同学合作, 但严禁抄袭
- 实验结果严禁造假, 我们会抽查部分同学的代码进行测试

## 提交

- 将报告(pdf), 源代码以及readme文件打包提交
- 命名格式: 学号\_姓名\_hw1.zip
- 截止时间: 2021.10.17