

## 期末 Proj 说明

### 任务：

标词性（2 个中文数据集）。

### 任务结果提交时间与形式：

- 1) 提交时间：7 月 4 日之前（实验报告 + 测试结果）。
- 2) 实验报告提交：包括你的方法，实验设置和步骤。将**训练数据**随机划分：70% 作为训练，30% 作为测试，请对测试结果评估（P/R/F），列出评估结果值，对方法和结果进行讨论和分析，形式与已经提交过的作业相同。
- 3) 测试结果提交：针对测试数据，直接提交测试结果，不需要自己评测（测试数据没有提供标准答案）。以**全部训练数据**训练模型，对**测试数据**进行测试，测试结果的表示与训练样本的表示格式完全相同（编码格式为 utf-16 little endian）。

### 对测试结果的说明：

- 1) 我们将统一对测试结果进行评估，以判断实验报告中结果的可信度，期末的成绩高度依赖测试结果。
- 2) 测试结果原则上与实验报告的结果没有太大的差异。
- 3) **如测试结果雷同（各指标完全一样），结果雷同的所有学生此次作业的期末成绩部分均记为 0 分。**

### 文件表示如下：

- 1) 文件的编码格式为 utf-16 little endian
- 2) 训练数据存放在 Train\_utf16 中
- 3) 测试数据放在 Test\_utf16 中（在 6 月 30 日提供）