

## “自然语言处理导论”小组作业要求

可选择候选或自选题目。

### 候选题目一：

CLUENER 细粒度命名实体识别：

<https://www.cluebenchmarks.com/introduce.html>。要求及评价方法、训练语料等均参考上述网站。

### 候选题目二：

作文打分，英语作文或汉语作文任选其一。训练语料分别为“作文打分语料（英语）.xlsx”“作文打分语料（汉语）.xlsx”。

评价标准：Spearman's  $\rho$ 、Pearson  $r$ 、RMSE、Cohen's  $\kappa$ 。

### 重要时间点：

5 月 12 日以前，小组确定题目，并提交给助教。自选题目的，同样需要在 5 月 12 日以前，和授课老师确认。

6 月 16 日提交文档、数据、代码。

文档至少包括以下几个部分：摘要、关键词、文献综述、重点难点、模型设计、实验结果、结论、参考文献、代码部署说明、小组分工说明及评分（组长为组员打分，满分为 100 分），可以结合代码一起使用 Jupyter。选择候选题目一的，需同时给出网站官方评价结果。

数据：预处理后的数据及模型。

代码：引用的非公开的包需连同小组编写的代码一同提交，涉及的文件均使用相对路径。代码需有必要的注释。如果代码无法执行，将影响最终得分。