# 《语言分析与机器翻译》课程实践

## 实践内容

课程实践可以在下面的题目中任选一个完成，或者自拟题目，但要求与本课程介绍的自然语言处理方法或模型相关。

1. **基础版：基于语言模型的古诗生成**

题目描述：参考实践环节代码，实现RNN语言模型，使用中国古诗数据集[1]进行训练，输入首字输出完整流畅的古诗。

1. **进阶版：基于LSTM-CRF模型的命名实体识别**

题目描述：实现LSTM-CRF模型，使用人民日报98版数据集[2]进行训练，输入文本输出其中实体及类别。

1. **挑战版：基于Transformer模型的机器翻译**

题目描述：实现Transformer模型，使用预处理好的IWSLT’14 De-En数据集[3]进行训练，输入德文输出对应的英文。

注：课程实践不限制所使用的深度学习框架，但禁止相互抄袭或照抄公开项目代码。

## 实践要求

1. 提交一份课程实践报告，命名格式为“学号-姓名”，详细描述实践内容、结果与心得体会，字数不少于3000。
2. 在GitHub网站上创建一个公开项目，将代码和课程报告提交至该项目，项目地址填写到群内的“[2021级语言分析与机器翻译课程实践项目汇总](https://docs.qq.com/sheet/DQUlFcVBScU5Tc2lT?tab=BB08J2)”腾讯文档中，注明学号和姓名。
3. 课程实践项目提交截止日期为12月1日。

## 成绩评定

1. 所有人必须提交课程报告，未提交课程报告直接取消成绩。
2. 在提交课程报告的基础上，项目完成度越高、效果越好或报告越完整则成绩越高。
3. 在评定成绩后，我们将选择3到4位表现突出的同学展示其课程实践内容（暂定远程会议，时间另行通知）。
4. 不得抄袭，如发现大面积雷同直接取消成绩。

## 参考资料

[1] <https://github.com/THUNLP-AIPoet/Datasets/tree/master/CCPC>

[2] <https://github.com/yaleimeng/NER_corpus_chinese>

[3] <https://git.io/JPK9N>

## 答疑

有任何问题请联系xiaotong@mail.neu.edu.cn（肖桐）或 huchinlp@gmail.com（胡驰）