Foundations of Natural Language Processing

Peking University, 2025

Assignment 4

For Graduating Students: Due Sunday, June 22, 2025 by 23:59
For Other Students: Due Monday, June 30, 2025 by 23:59

1. Directions

PLEASE read the instructions to ensure you receive full credit on your homework.

LATE SUBMISSION POLICY (Not applicable to graduating students)

- Late homework will have 5% deducted from the final grade for each day late.

- No submission will be accepted after July 7, one week after the due date. Your homework submission time will be based on the time of your last submission to Course.

- Please be aware that all late submissions, regardless of the reason, will be penalized according to the late submission policy. While we understand that unexpected events can occur, it is your responsibility to plan accordingly and submit your homework on time.

2. General Descriptions

In this lab, you are going to complete **one of the following** NLP projects according to your interest. See the detailed instructions for each project in

their respective PDFs.

Project 1: Machine Translation for Low-Resource Languages

Project 2: Multi-Defendant Legal Judgment Prediction

In general, you should follow the following steps:

Step 1: Choose your project

First, choose one of the three projects list above and join the Kaggle leaderboard for this task by the following links.

Project 1:

Sub-Task 1:

https://www.kaggle.com/t/5b442ae0a424454d862f43b5b8c5b384

Sub-Task 2:

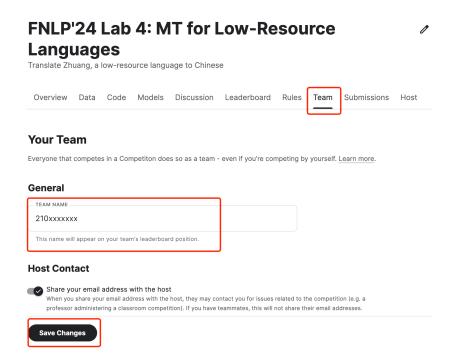https://www.kaggle.com/t/cf84fc6677d94c9f83940a81da0da0f9

Project 2:

Sub-Task 1:

https://www.kaggle.com/t/853b5d7dc4eb490e834745d900df9e3d

Sub-Task 2:

https://www.kaggle.com/t/92afa88d912a478e8cb70223a4427c2a

Please change your Team Name to your student ID. Otherwise, we cannot identify your scores on the leaderboard and you will get zero grade.

FNLP'24 Lab 4: MT for Low-Resource Languages ✎

Translate Zhuang, a low-resource language to Chinese

Overview Data Code Models Discussion Leaderboard Rules **Team** Submissions Host

Your Team

Everyone that competes in a Competiton does so as a team - even if you're competing by yourself. Learn more.

General

TEAM NAME

210xxxxxxx

This name will appear on your team's leaderboard position.

Host Contact

⬤ Share your email address with the host
When you share your email address with the host, they may contact you for issues related to the competition (e.g. a professor administering a classroom competition). If you have teammates, this will not share their email addresses.

Save Changes

You can change your choice of project before the deadline as you wish. However, make sure that you clearly indicate your final choice in your report. If you write about more than one project in your report, we will only grade the first one and ignore the rest.

Step 2: Carry out your project

You should carefully read the instructions of your chosen project. Here is some general information that you should be informed of.

Submission Trials:

Please actively use our leaderboard system to evaluate your methods. **Note that you can only submit your prediction results up to five times every 24 hours.** Do not delay your project to the last few days!

LLM APIs:

According to each project description, you might conduct some

experiments with LLMs. **To ensure a fair comparison between students, please use the qwen-max API from Aliyun** (https://help.aliyun.com/zh/model-studio/models?spm=a2c4g.11186623.0.0.78d84823osmlX0#cfc131abafghw) **for your experiments with LLM APIs.** Aliyun provides each new account with 1M tokens for free, which is sufficient for the project.

Step 3: Submit your project

<u>Submission Deadline:</u>

By June 30, 2025, you should submit all the material required in the instructions to course.pku.edu.cn.

For graduating students, your deadline is earlier (June 22). Don't miss the deadline, or it would affect your graduation. **The late submission policy does not apply to the graduating students.**

<u>Grading Criterion:</u>

We will grade your project based on the following criterion:

1. Report (80%): In the report, you should

   a) describe your method clearly;

   b) report the implementation details of your method;

   c) evaluate your method on the leaderboard;

   d) compare your method with a suitable baseline (such as a simple zero-shot baseline for LLMs) to evidence the effectiveness of your method;

   e) analyze the experiment results in depth.

2. Implementation (5%): We will check whether you have submitted all the required material. The submitted code should meet the requirement of reproducibility.

3. Leaderboard (15%): Part of the grade will be determined by your ranking on the leaderboard.

3. Contact

If you have any general questions about Assignment 4, please contact zhangch@pku.edu.cn.

If you have questions for a specific project, please contact the following TAs:

Project 1: zhangch@pku.edu.cn linjiuheng@stu.pku.edu.cn

Project 2: luokangcheng@stu.pku.edu.cn ziruiwu@pku.edu.cn

Good luck with your final project!