

作业二：实现HMM中文分词和BPE英文分词

1. 作业要求

基于提供的Python文件/Jupyter Notebook文件，以代码填空的形式，实现HMM中文分词和BPE英文分词。填空完毕后，需展示代码中相应测试部分的输出结果。本次作业，HMM中文分词和BPE英文分词各9分，共计18分。

提示：只需填写代码中TODO标记的空缺位置即可，具体的代码说明和评分细则详见提供的脚本文件。

2. 提交方式

以下两种方式选择其一提交至CANVAS平台即可

- 直接提交带有已补全的代码及运行结果的Jupyter Notebook文件，文件中要求对所填代码作必要的说明。
- 提交已补全的两个Python文件和实验报告，报告中要求对代码作必要的说明并展示运行结果。

3. 文件说明

- data 含有指定的训练与测试语料
- hmm_parameters.pkl hmm的参数文件
- hmm_bpe.ipynb 提供的Jupyter Notebook文件
- hmm.py 提供的Python文件（任务一HMM相关）
- bpe.py 提供的Python文件（任务二BPE相关）