

课后作业 HW07

手写或打印一页A4纸/题

课堂提交

作业编号 7
满分10分

Q1 (5分) 实现 Bi-Gram PPL 计算

- 下载 [demo](#), 查看 `nlp_bigram_ppl_TODO.py`
- 给定语料库: I play tennis. I like Chinese friends. I talk with Chinese student. I play with tennis friends. I have friends who like tennis.
- 分别计算 “I play with Chinese friends” 与 “Chinese friends who like tennis” Bi-Gram PPL. 比较大小, 说明原因。
- 代码中用了 NLTK 分词器, 查阅资料并简要说明分词原理

Q2 (5分) Hamlet's Question

- 查看给定的 William Shakespeare 文集
- 阅读readme，下载 LLAMA-3 分词器。
- 计算 "To be, or not to be: that is the question" 困惑值PPL
- 实现 nlp_tokenizer_shakespeare_TODO.py