知识表示与推理课程

命名实体识别实验报告

姓名：徐柯炎

学号：2021110683

日期：2024.5.28

1. **基于ltp的命名实体识别**
2. 简述命名实体识别相关概念。）

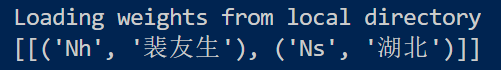
命名实体识别(Named Entity Recognition，简称NER)是自然语言处理(NLP)中的一项重要任务，旨在从文本中识别并分类具有特定意义的实体，如人名、地名、组织机构、日期、时间、百分比、货币等。NER的目标是识别文本中与特定类别相关的命名实体，并为这些实体分配预定义的类别标签。

NER系统通常采用监督学习方法，使用带有标签的训练数据进行模型训练。训练数据包含了文本句子和对应的实体标签，模型通过学习这些示例来预测新的文本中的实体和它们的类别。

1. 更改例句，使用ltp库完成命名实体识别，分析对应识别结果。）

这里采用的例句为：“裴友生，男，汉族，湖北蕲春人，1957年12月出生，大专学历。”

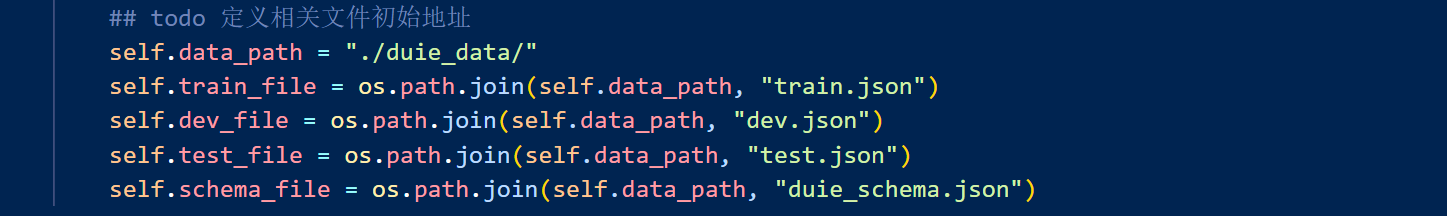
结果如下：

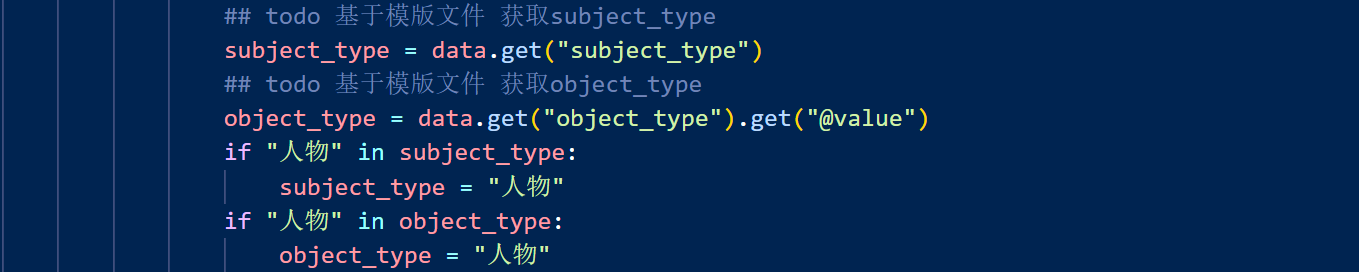


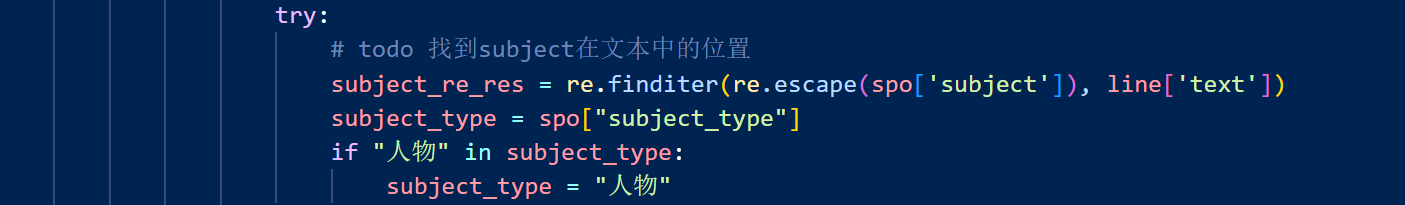
分析识别结果：“裴友生”是人名，“湖北”是地名，分析结果基本正确。

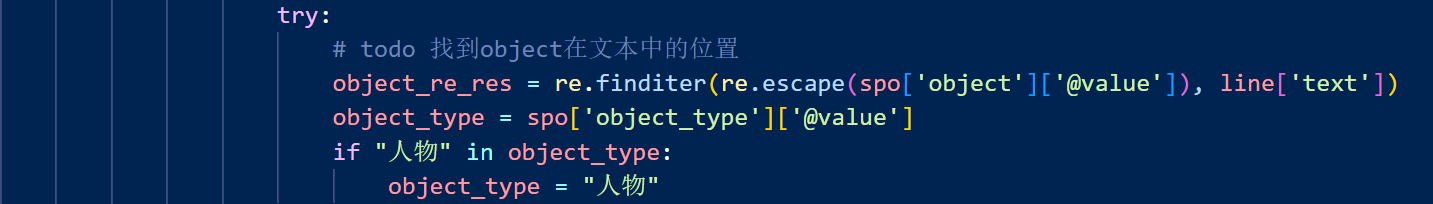
1. **基于bert的命名实体识别**
2. 补全各部分代码，完成手册中三部分实验内容，在报告中以截图方式展示补全代码。）

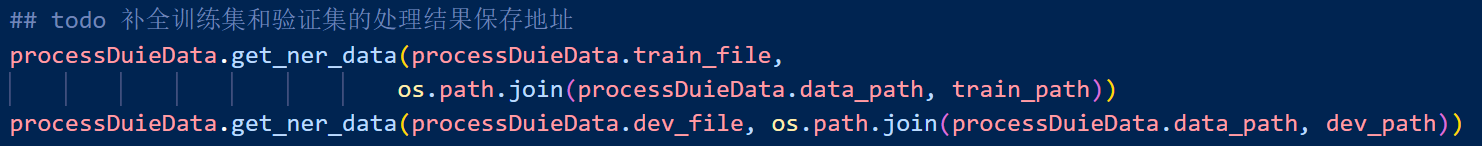
Process.py



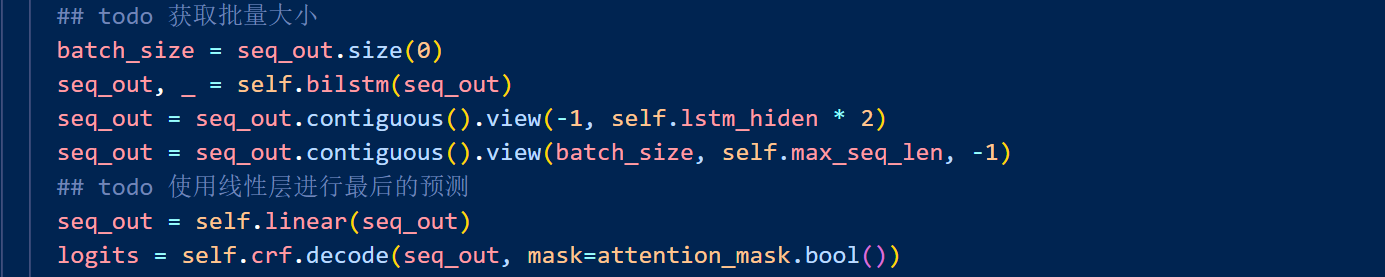




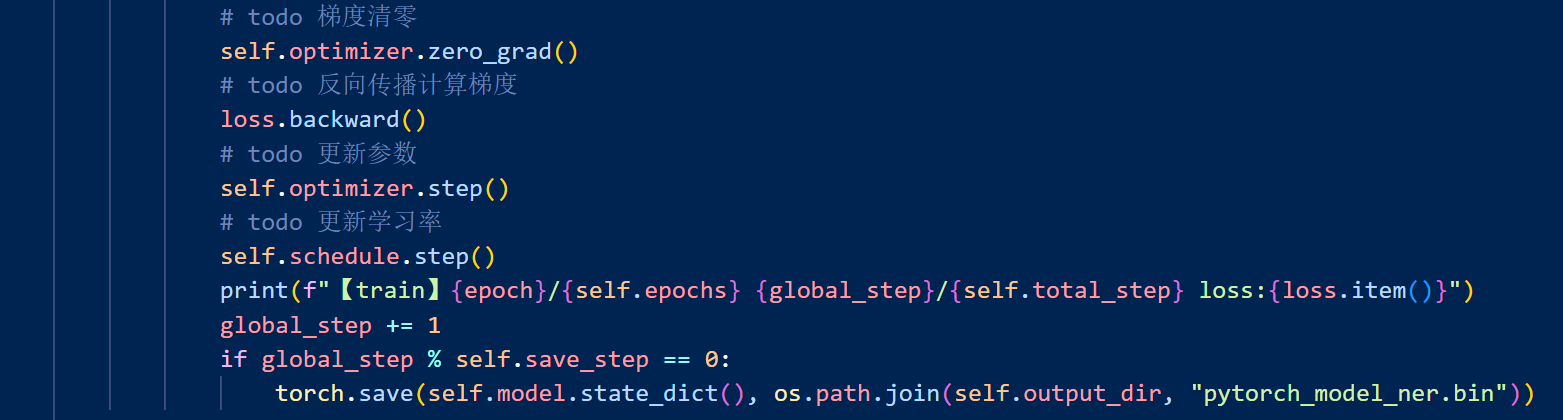


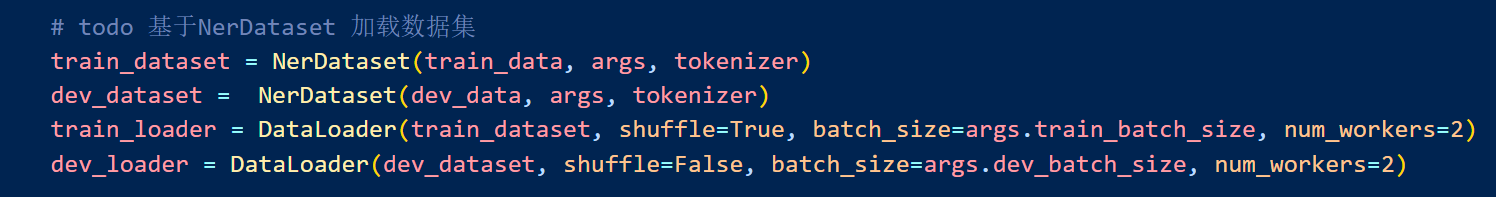


model.py

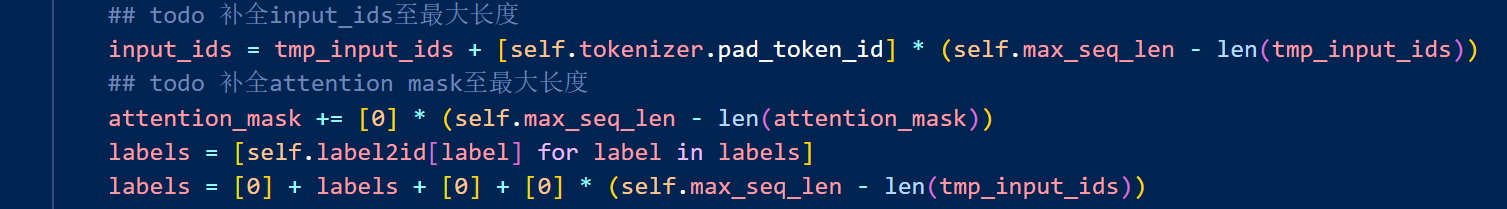


main.py





data\_loader.py



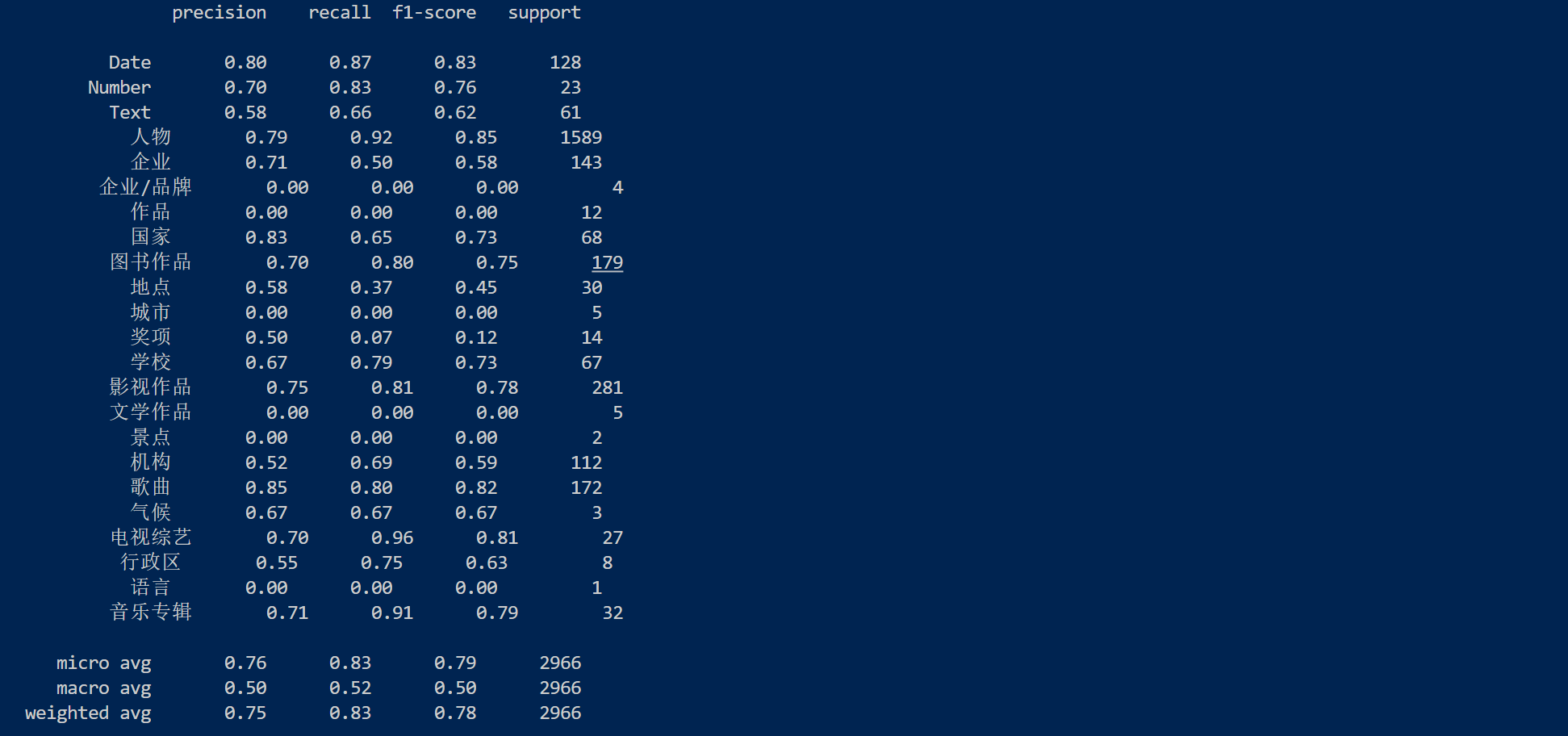
1. 对于训练部分给出训练成功截图，并分析模型训练代码逻辑。）

训练代码逻辑如下：

1. 导入预先设置的超参数，并将超参数写入ner\_args.json中，方便后续调用；
2. 导入tokenizer并且指定使用的模型，设置使用的device（不能用cuda就使用cpu运算）；
3. 准备使用的数据集：读入train.json和dev.json，分别用于训练和测试，然后使用NerDataset类加载数据集，并通过DataLoader导入；
4. 准备optimizer和训练策略（在训练的过程中调整学习率）：
5. 然后通过Trainer类来加载训练模型；
6. 开始训练：计算模型输出，然后反向传播并更新参数和学习率，计算损失函数；
7. 最后测试模型。

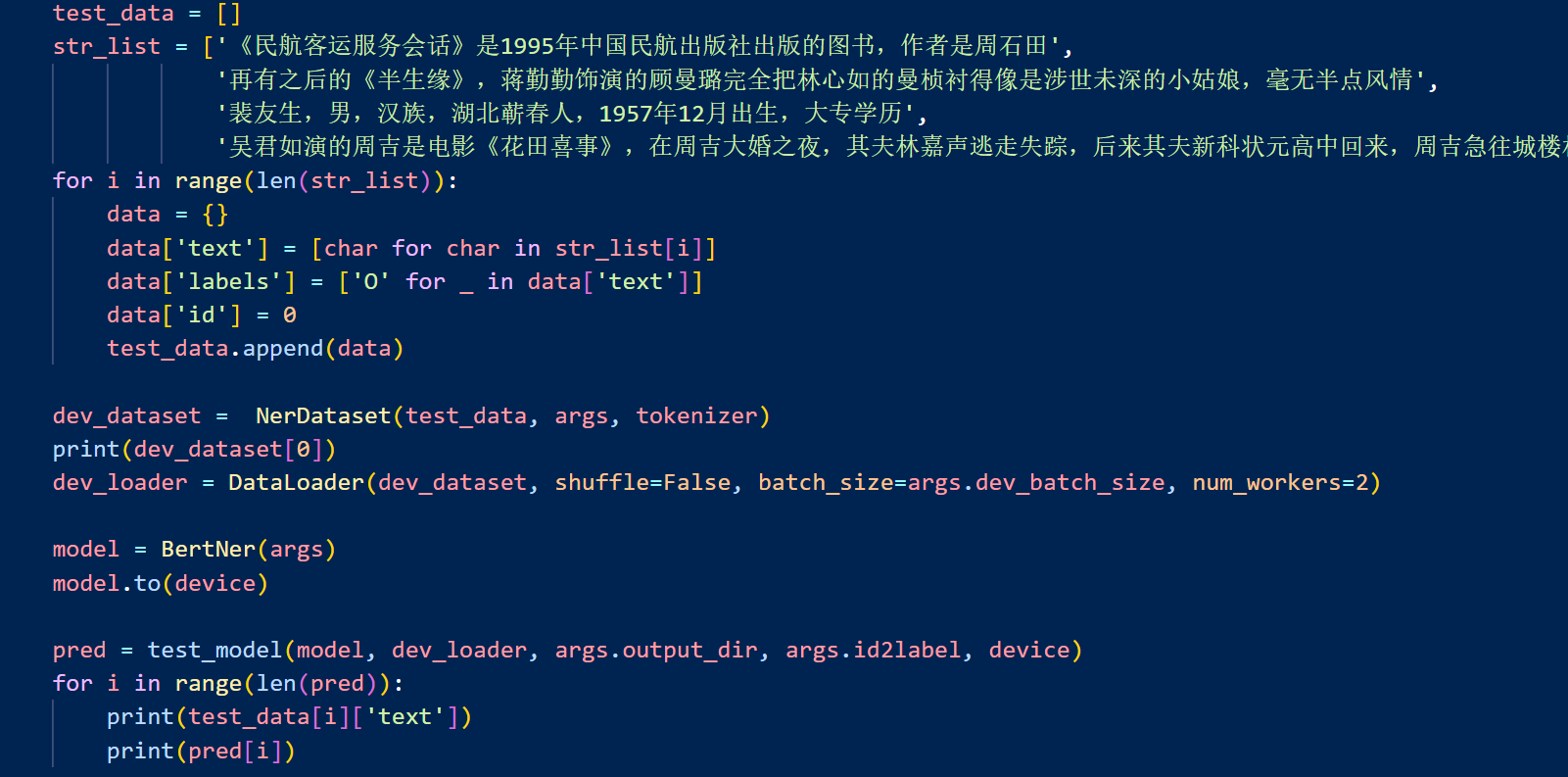
训练成功截图：



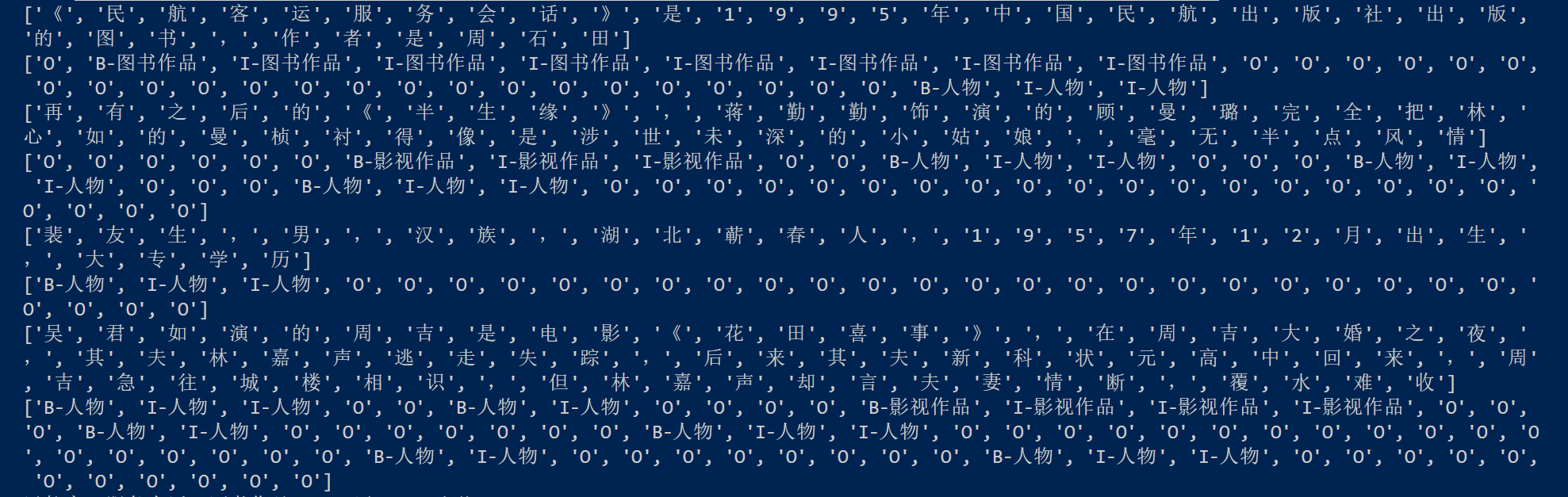


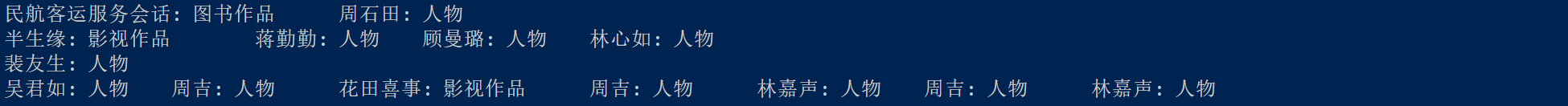
1. 提供预测部分代码，并给出手册中例句的命名实体识别结果。）

部分预测代码如下：



手册中例句的命名实体识别结果如下：





1. **基于大模型的命名实体识别**
2. 至少采用两种大模型完成例句的命名实体识别，并按照json格式展示实验结果）

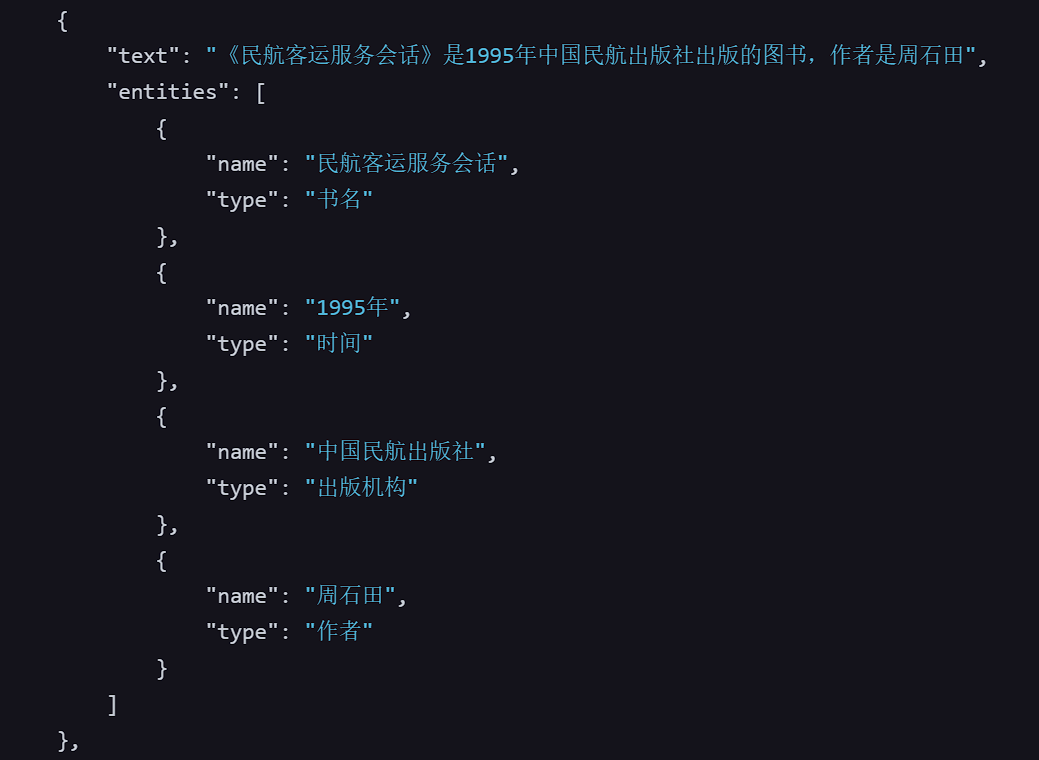
以下是我采用两种大模型完成例句的命名实体识别的prompt：

下面将给你几段文本数据，请你对此文本进行命名实体识别，并按照json格式输出结果。数据如下：“['《民航客运服务会话》是1995年中国民航出版社出版的图书，作者是周石田', '再有之后的《半生缘》，蒋勤勤饰演的顾曼璐完全把林心如的曼桢衬得像是涉世未深的小姑娘，毫无半点风情', '裴友生，男，汉族，湖北蕲春人，1957年12月出生，大专学历', '吴君如演的周吉是电影《花田喜事》，在周吉大婚之夜，其夫林嘉声逃走失踪，后来其夫新科状元高中回来，周吉急往城楼相识，但林嘉声却言夫妻情断，覆水难收']”

展示实验结果：

例句1：“《民航客运服务会话》是1995年中国民航出版社出版的图书，作者是周石田”

文心一言：

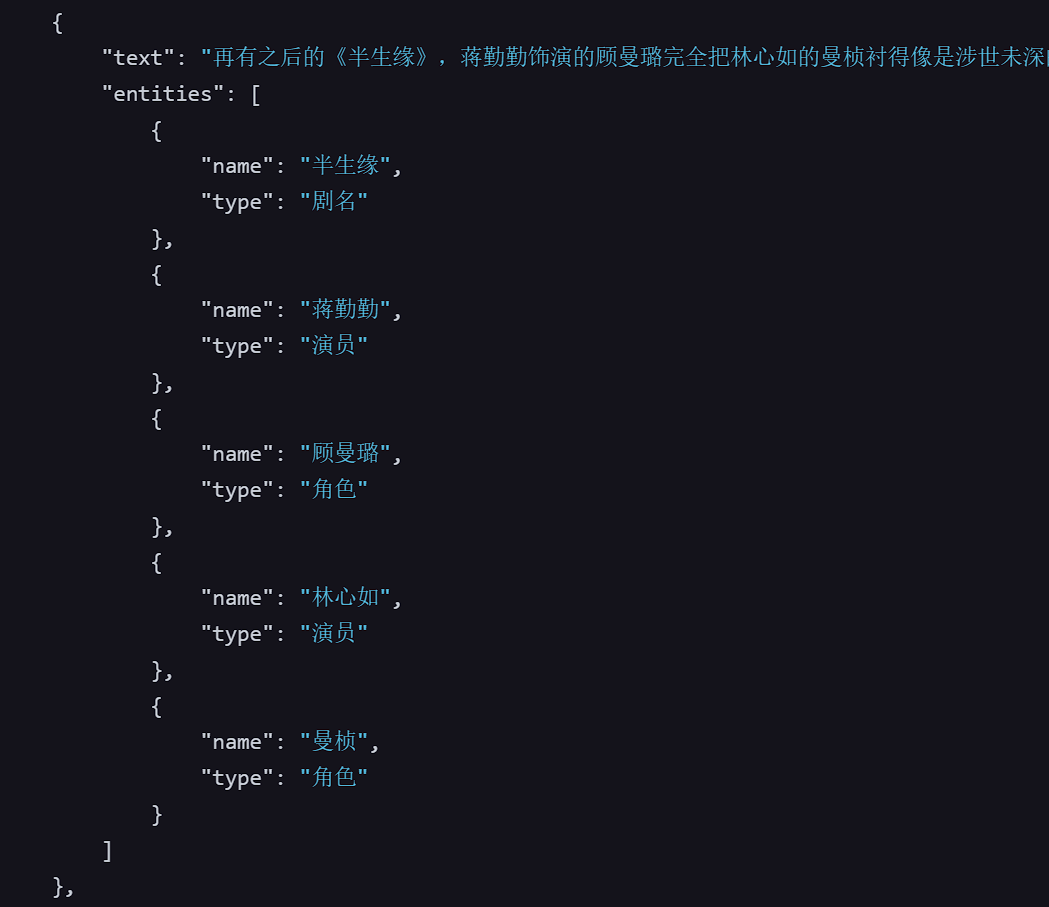


通义千问：



例句2：“再有之后的《半生缘》，蒋勤勤饰演的顾曼璐完全把林心如的曼桢衬得像是涉世未深的小姑娘，毫无半点风情”

文心一言：



通义千问：



例句3：“裴友生，男，汉族，湖北蕲春人，1957年12月出生，大专学历”

文心一言：



通义千问：



例句4：“吴君如演的周吉是电影《花田喜事》，在周吉大婚之夜，其夫林嘉声逃走失踪，后来其夫新科状元高中回来，周吉急往城楼相识，但林嘉声却言夫妻情断，覆水难收”

文心一言：



通义千问：

