1. 实验题目：基于HMM的拼音转汉字程序

（1）实验目的

理解、掌握隐马尔可夫模型，N元语法等自然语言处理的基本思想、算法，并将其应用于从汉语拼音到汉字的自动转换过程。

假定：拼音串中已经用空格进行了分隔，如“wo ai wo jia”

（2）实验项目内容

1）对训练语料及相关资源进行预处理；

2）通过学习算法，训练HMM模型；

3）利用HMM模型和维特比算法，实现从任意拼音到汉字的自动转换。

4）利用给定测试集，评价上述程序的转换准确率。

2、实验题目：基于ModelArts平台的文本分类实验

（1）实验目的

理解、掌握深度学习模型CNN和RNN的结构和华为MindSpore框架，掌握使用ModelArts模型构建、训练、分类与性能评估方法。

（2）实验项目内容

1）练习使用ModelArts开发平台，包括开发流程、对象存储服务、自定义模型和预置模型加载、运行等。

2）使用MindSpore或pytorch或tensorflow框架，开发一基于CNN和RNN的文本分类模型，对文本所属类别进行分类（情感级性为正向或负向）。

3）在ModelArts平台上，对所开发的分类模型进行测试和优化，对比有dropout和无dropout、使用单层和多层卷积池化层(对CNN)、使用单向和双向LSTM层（对RNN）后的结果。

注意：提供的代码材料主要是基于CNN模型的，大家可以在所提供的材料基础上修改，使其也适用于RNN模型。

3、实验题目：基于ModelArts平台的文本生成实验

（1）实验目的

通过本实验，学生可以学习BERT模型、LSTM模型在序列生成中的使用，文本的预处理操作。熟悉神经网络的结构，文本生成的过程以及网络模型的训练和预测，掌握MindSpore的相关操作等。

（2）实验项目内容

本实验的基本内容就是使用序列生成模型，如BERT、LSTM、RNN等完成诗歌自动生成的功能，训练完成后通过加载模型和网络结构，生成测试诗句。

1) 华为云ModelArts环境搭建，下载并分析数据集；

2) 对训练数据集中的诗词进行预处理，如对多首诗词进行拆分、分割题目、提取内容、生成训练样本片段等。

实验3

3）使用BERT、LSTM、RNN中的任意一种模型，设计并构建一古诗词生成网络模型，训练并输出模型的Loss变化曲线，保存模型结构与参数。

4) 加载已训练模型，通过指定参数（如5言，7言）生成测试古诗。

实验4

5）通过输入每句诗中的头一个字和参数（如5言，7言），生成测试藏头诗。

6) 在测试集上计算生成模型的rouge-1，rouge-2性能指标（选做）。