**一、实验说明**

1. 在8核i7处理器下，代码大约要运行 1min30s左右，请耐心等待。

2. 实验使用HMM（隐马尔科夫模型）进行中文分词。测试集的分词结果存放在result.txt中。

3. 没有给标准输出结果啊，小改改，没法输出precision、recall、F1啊…难受…如果想让我完成未竟的事业，请cqy我一下，将标准输出给我，我做一下哈。

4. 请将所需的数据文件：CTBtestingset.txt、CTBtrainingset.txt放在同级目录下以保证代码正确运行

5. 祝身体健康，每天开心愉快！

**二、实验代码**

代码建模最基础的HMM，使用HMMSegger继承了HMM以具体完成中文分词任务。

HMM模型部分：

L4 – L23 HMM模型架构。

L40 – L43 加载数据代码。

L45 – L61 有监督学习的训练代码。

L63 – L81 预测前的一步，计算出相应概率。

L83 – L114 Viterbi算法代码，用来预测结果。

功能函数部分：

L120 – L131 为训练集标注标签(B,E,M,S)。

L133 – L164 根据预测结果标签将中文分词。

HMMSegger模型部分：

L175 – L199 Segger训练代码。

L201 – L206 Segger预测并分词代码。

**三、实验结果**

代码运行结束后，结果存放在result.txt中，部分截图如下：

