

# 拼音输入法

---

计82 洪昊昀

## 1. 运行拼音输入法程序

python 版本: 3.8.2

运行环境: Win 10, Pycharm Community Edition 2019.3.4

`pinyin.py` 不依赖外置的包

请直接在 `src` 目录下按如下方式运行 `pinyin.py`, 否则会出现路径问题:

```
python pinyin.py ../data/input.txt ../data/output.txt
```

会在屏幕上输出类似如下内容:

```
E:\洪昊昀_拼音输入法\src>python pinyin.py ../data/input.txt ../data/output.txt
['qing', 'hua', 'da', 'xue', 'ji', 'suan', 'ji', 'xi']
清华大学计算机系
['wo', 'shang', 'xue', 'qu', 'le']
我上学去了
['jin', 'tian', 'hui', 'jia', 'bi', 'jiao', 'wan']
今天回家比较晚
['liang', 'hui', 'zai', 'bei', 'jing', 'zhao', 'kai']
两会在北京召开
```

答案会被添加到 `../data/output.txt` 文件中

## 2. 项目结构

### 2.1 readme

### 2.2 data

#### 2.2.1 input.txt (会被替换的输入文件)

#### 2.2.2 output.txt (会被替换的输出文件)

### 2.3 src

#### 2.3.1 pinyin.py

为拼音转换启动代码。

### 2.3.2 model\_2\_3\_4\_gram.py

转换模型

### 2.3.3 counts 和 hanzi\_pinyin

存储模型参数

语料、处理语料的文件、产生模型过程前的代码及它们的**readme**请见清华云盘网址：<https://cloud.tsinghua.edu.cn/d/d96d9124ecfd42a4a671/>