

源代码说明

data

目录保存了模型所需要的数据和测试数据

zi.json 保存了所有的汉字

homoform.txt 保存了字形相似的字

ci.json 保存了词和词频

data.txt 为人民日报语料库(未带词性)

test.txt 和 real.txt 是测试数据， test.txt 经人为修改得到，是自己制作的测试集

getdata

zi.py 是使用爬虫获取汉字数据的代码

ci.py 是使用爬虫获取词数据的代码

process.py 是生成测试集的代码

main

main.py 为主程序的入口，调用的其他文件如下

Tree.py 是保存词所用的字典树的数据结构

load_data.py 是加载所需数据的代码

AutoCorrect.py 是对句子进行纠正的源代码，并提供句子纠正的接口

eval.py 是测试模型准确率的源代码

可执行程序

只能在当前目录运行!!!

main.exe 是生成的可执行程序，直接在当前目录下运行即可(移动到其他目录会导致找不到所需文件而无法运行)

由 pyinstaller 生成