

用户手册

一、文件准备

1. 您需要将 wordsep.cpp 编译为可执行文件 wordsep.exe，或者直接使用文件夹中的 wordsep.exe。

2. 您需要确保文件夹中的 dict.txt 是以 GBK 格式进行保存的（查看方式：用记事本打开 dict.txt，查看右下角是否为“ANSI”，如果不是的话请用记事本另存为“ANSI”格式，或参考这篇文章：https://blog.csdn.net/weixin_44578029/article/details/109205699），自带 dict.txt 是以 GBK 格式保存的。

3. 文件夹“词典文件”中有两种类型的文件，其中“dict_small”是一些较常用的词，“dict_large”包含更多的词语，包括一些中英文混杂的词语，如“C 盘”。使用时将希望使用的词典取出，放入 wordsep.exe 同级的文件夹即可。初始时 wordsep.exe 同级目录中 dict.txt 为 dict_small 中文件。

二、使用方法

1. 本程序的执行文件为 wordsep.exe，并且需要词典文件 dict.txt。需将词典文件放在程序的运行目录下。

2. 双击 wordsep.exe 进入演示程序后，程序会开始读取 dict.txt 中的内容，您可以查看进度条了解词典加载进度。

3. 当词典加载完成后，程序会提示您输入一串需要分词的字串。您需要输入希望进行分词的字符串。

4. 之后程序会提示“请输入希望获得的分词类型数（n 最短分词）”，此时您需要输入希望获得的路径种类数。

5. 片刻后，程序会输出分词结果。用户界面如下所示：



```
e:\vscode_c\wordsep.exe
活动代码页: 936
词典构建完成!
请输入待分词的中文语句:
>>>我是中国人民大学高瓴人工智能学院的一名大二本科生,学人工智能,就来高瓴人工智能学院!
请输入希望获得的分词类型数(n最短分词):
>>>3
序号 分词结果
1 我/是/中国人民大学/高/瓴/人工智能/学院/的/一名/大二/本科生/, /学/人工智能/, /就/来/高/瓴/人工智能/学院/!
2 我/是/中国人民大学/高/瓴/人工智能/学院/的/一名/大二/本科生/, /学/人工智能/, /就/来/高/瓴/人工智能/学院/!
3 我/是/中国人民大学/高/瓴/人工智能/学院/的/一名/大二/本科生/, /学/人工智能/, /就/来/高/瓴/人工智能/学院/!
是否继续?[y/n]
>>>y
请输入待分词的中文语句:
>>>我爱数据结构与算法!
请输入希望获得的分词类型数(n最短分词):
>>>4
序号 分词结果
1 我/爱/数据/结构/与/算法/!
2 我/爱/数据/结构/与/算法/!
3 我/爱/数据/结构/与/算/法/!
4 我/爱/数据/结构/与/算/法/!
是否继续?[y/n]
>>>
```

6. 当您希望获得的路径方法数多于实际的路径时，程序会提示您路径最多有多少种。

7. 之后，程序会提示“是否继续？[y/n]”。如果您希望接着进行分词，请输入“y”，如果您不想继续分词，请输入“n”。请注意，必须输入“y”或者“n”中的一个，否则程序会提示“非法输入！”并继续询问您“是否继续？[y/n]”。

三、注意事项

1. 本程序的词典表 dict.txt 是可以自定义的，您只需要保证自定义词语小于 25 个字、

每一行的格式为“词[空格]词频[空格]（可选）词性”，以及保存格式为 GBK 即可。

2. 本程序支持的**最大字符串长度为 100**，也就是说，您需要确保您输入的字符串长度小于 100 个字，请注意，英文符号以及空格等**标点符号也算作一个字**。

3. 本程序支持查询的**最大路径数为 30**，也就是说，程序无法返回超过 30 种分词结果，因此，在程序提示“请输入希望获得的分词类型数(n 最短分词)”时，您最好需要输入一个不超过 20 的数，不过，也不要过于担心，当路径数超过 30 时本程序会提示您并显示前 30 种解决方案。

4. 本程序支持对于**混有英文的词语（如“c 盘”）的分词**，只要将这些词添加到 dict.txt 中即可。

最后，祝您生活愉快！