**英语单词词形还原**

孟磊 MF1833048

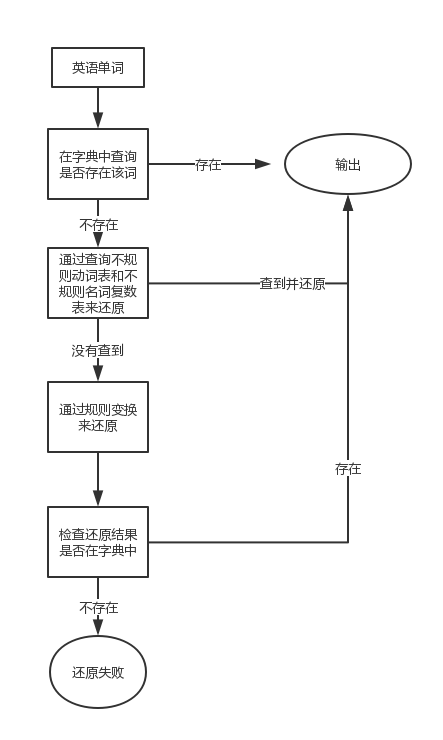
**程序运行方式: 目录下的 token文件为程序入口，英语单词为参数，即 ./token 单词**

**运行环境为 python3.7**

1. **任务描述**

设计并实现程序，把输入的英文单词还原成其原形。

1. **技术路线**



规则变换用到的规则和规则的顺序如下：

(1)\*ves --> \*f/\*fe (8)\*??ing --> \*?

(2)\*ies --> \*y (9)\*ying --> \*ie

(3)\*es --> \* (10)\*ing --> \*/\*e

(4)\*s --> \* (11)\*??ed --> \*?

(5)\*ies --> \*y (12)\*ied --> \*y

(6)\*es --> \* (13)\*ed --> \*/\*e

(7) \*s --> \*

1. **用到的数据**
2. 英文单词词典数据：

来源为：<http://nlp.nju.edu.cn/MT_Lecture/dic_ec.rar>，整理为目录下的dic\_ec.txt

（2）不规则动词数据：

来源为：

<https://baike.baidu.com/item/%E8%8B%B1%E8%AF%AD%E4%B8%8D%E8%A7%84%E5%88%99%E5%8A%A8%E8%AF%8D%E8%A1%A8/1619648?fr=aladdin>

整理为目录下的：IrregularPluralNouns

（3）不规则名词数据：

来源为：<https://wenku.baidu.com/view/bee71a621ed9ad51f01df2d3.html>

整理为目录下的：IrregularVerbList

1. **遇到的问题以及解决方案**
2. 有些词，比如一些词的ing形式、found，它们本身就存在于词典文件中，是一个独立的词，所以会在词典中匹配后就返回结果，并不会进行词形还原。所以在程序中做了修改，在词典中匹配到后，还会继续进行词形还原，这两个结果会一起输出。
3. **性能评价**

时间复杂度较好。正确率取决于数据集的完备性。本次实验中，还原正确率很好。