主要功能：

- 中文分词 s.words
- 词性标注 s.tags
- 情感分析 s.sentiments
- 文本分类
- 转换成拼音 s.pinyin
- 繁体转简体 s.han
- 提取文本关键词 s.keywords(number)
- 提取文本摘要 s.summary(number)

classification
normal ->转换成拼音、繁体转简体、分句
seg -> 中文分词
sentiment ->情感分析
sim -> 文本相似度
summary ->提取文本关键词、提取文本摘要
tag ->词性标注
utils ->一些辅助函数

准备工作
http://www.runoob.com/python/python-object.html

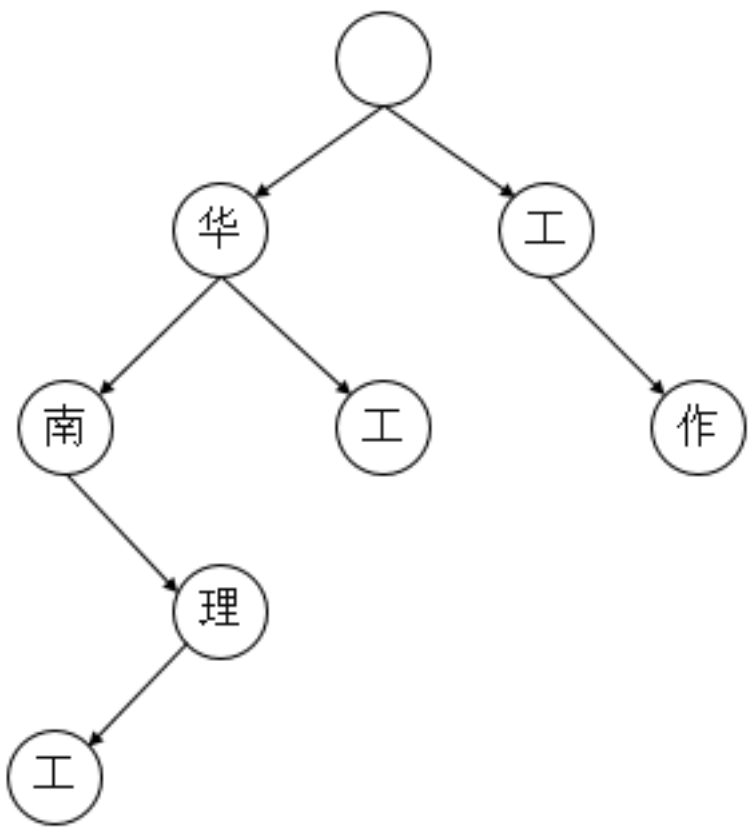
def
https://www.cnblogs.com/qlshine/p/6049457.html

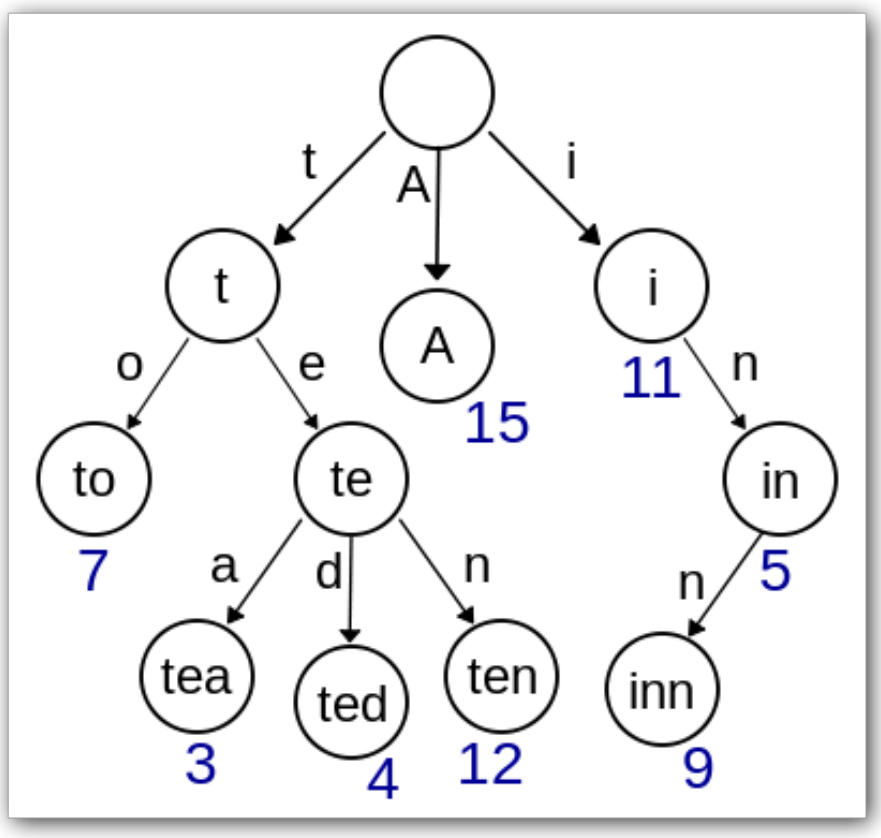
re模块：提供了正则表达式匹配操作

https://blog.csdn.net/liyahui_3163/article/details/78434157
re.compile(pattern[,flags])根据包含正则表达式的字符串创建模式对象，flags是匹配模式
可以实现更有效率的匹配。在直接使用字符串表示的正则表达式进行search,match和findall操作时，python会将字符串转换为正则表达式对象。而使用compile完成一次转换之后，在每次使用模式的时候就不用重复转换。
re.match(pattern, string, flags = 0)
如果在*字符串*的开头的零个或更多字符匹配正则表达式*模式*，将返回相应的MatchObject实例。返回None则该字符串中与模式不匹配；请注意这是不同于零长度匹配。即使在多行模式下， re.match()将只匹配字符串的开头，而不是在每个行的开头。

trie树

字典树，利用字符串的公共前缀来节约存储空间
https://blog.csdn.net/handsomekang/article/details/41446319
https://blog.csdn.net/v_july_v/article/details/6897097





1. 插入新单词
2. 查找单词
3. 输出

bm25

https://blog.csdn.net/Oscar6280868/article/details/81288772
定义：常见的用来计算query和文章相关度的算法。
原理：将需要计算的query分词成w1，w2,….，wn，然后求出每一个词和文章的相关度，最后将这些相关度进行累加。

textrank
TnT
情感分析
https://blog.csdn.net/google19890102/article/details/80091502
https://www.cnblogs.com/zz22--/p/9351346.html