Tokenizer

类成员变量

lock

dictionary

FREQ: 词典{} 前缀词典

total：int 类型

user\_word\_tag\_tab: dict 类型

tmp\_dir: 工作目录

initialized: bool 类型

cache\_file:string 类型

里面存着前缀词典和所有词的频率和。

类成员函数

\_\_init\_\_

\_\_repr\_\_

gen\_pfdict(self, file\_type):

输入前缀词典文件类型（file type in file object）

输出词典和所有词的频率总数

initialize(self, dictionary=None):

初始化tokenizer，输入词典文件名，加载前缀词典并且存储词典核所有词的频率总数

check\_initialized(self):

检查tokenizer是否初始化，没有初始的话就初始。

get\_DAG(self, sentence):

输入句子，输出有向无环图DAG。其中DAG的key是词在句中index，value是list类型。包括所有以key为起点，index\_为终点，其中index\_满足sentence[key, index\_+1]存在于前缀词典中。

calc(self, sentence, DAG, route):

输入有向无环图