#!/usr/bin/python3

# -\*- coding:utf-8 -\*-

from whoosh.qparser import QueryParser

from whoosh.index import create\_in

from whoosh.index import open\_dir

from whoosh.fields import \*

from jieba.analyse import ChineseAnalyzer

#from get\_comment import SQL

from whoosh.sorting import FieldFacet

import io, os, sys, time

sys.stdout = io.TextIOWrapper(sys.stdout.buffer,encoding='utf-8')

# 导入中文分词工具

analyser = ChineseAnalyzer()

# 创建索引结构: 没有结构，就是很多文本文件，一行一行的，总之很多行

schema = Schema(full\_line=TEXT(stored=True, analyzer=analyser))

# 数据和索引所在目录，以及索引名称

ix = create\_in("shegongku\_idx", schema=schema, indexname='allin1line')

# 返回root下的文件列表（不包含子目录）

def traverseFile(root):

flist = []

for f in os.listdir(root):

f\_path = os.path.join(root, f)

if os.path.isfile(f\_path):

flist.append(f\_path)

print(f\_path)

else:

flist += traverseFile(f\_path)

return flist

# 处理数据文件，每个文件每一行，加进去

writer = ix.writer()

for fn in traverseFile("shegongku\_db"):

with open(fn, 'r', encoding='utf-8') as f:

print(fn, "...")

lines=0

while True:

line1 = f.readline()

if line1:

writer.add\_document(full\_line=line1)

lines+=1;

else:

break

print(fn, lines, "added")

writer.commit()

print("index finished")

# 以上为建立索引的过程

index1 = open\_dir("shegongku\_idx", indexname='allin1line')

parser1 = QueryParser("full\_line", index1.schema)

while True:

with index1.searcher() as searcher:

print("pls input what u want to search:")

key = input()

myquery = parser1.parse(key)

resultss = searcher.search(myquery, limit=2000)

#print(type(resultss))

for result1 in resultss:

d1=dict(result1)['full\_line']

print(d1)