

```

pip install scikit-learn

Looking in indexes: https://pypi.org/simple, https://us-python.pkg.dev/colab-wheels/public/simple/
Requirement already satisfied: scikit-learn in /usr/local/lib/python3.9/dist-packages (1.2.2)
Requirement already satisfied: threadpoolctl>=2.0.0 in /usr/local/lib/python3.9/dist-packages (from scikit-learn) (3.1.0)
Requirement already satisfied: numpy>=1.17.3 in /usr/local/lib/python3.9/dist-packages (from scikit-learn) (1.22.4)
Requirement already satisfied: scipy>=1.3.2 in /usr/local/lib/python3.9/dist-packages (from scikit-learn) (1.10.1)
Requirement already satisfied: joblib>=1.1.1 in /usr/local/lib/python3.9/dist-packages (from scikit-learn) (1.1.1)

import jieba
import json
from sklearn import metrics

with open("movie_all_info.json") as movie:
    movie_info=json.load(movie)
with open("stopword.txt","r",encoding="utf8") as f:
    stopword=f.read()

class Appearance:
    def __init__(self,docId,frequency):
        self.docId=docId
        self.frequency=frequency
    def __repr__(self):
        return str(self.__dict__)

class Database:
    def __init__(self):
        self.db=dict()
    def __repr__(self):
        return str(self.db)
    def get(self,id):
        return self.db.get(id,None)
    def add(self,document):
        return self.db.update({document['id']:document})
    def remove(self,document):
        return self.db.pop({document['id']:document})

class InvertedIndex:
    def __init__(self,db):
        self.index=dict()
        self.db=db
    def __repr__(self):
        return str(self.index)
    def index_doc(self,document):
        terms=[t for t in jieba.cut(document["intro"],cut_all=False) if t not in stopword]
        appearances_dic=dict()
        for term in terms:
            # print(term)
            term_frequency=appearances_dic[term].frequency if term in appearances_dic else 0
            appearances_dic[term]=Appearance(document['id'],term_frequency+1)
            # print(appearances_dic[term])
        #把invert update
        update_dic={key:[appearance] if key not in self.index
                    else self.index[key]+[appearance]
                    for (key,appearance) in appearances_dic.items()}
        self.index.update(update_dic)
        self.db.add(document)

        return document
    def lookup_query(self,query):
        return{term:self.index[term] for term in query.split(" ") if term in self.index}

db = Database()
index = InvertedIndex(db)
for i in range(len(movie_info)):
    doc=dict(id=movie_info[i]["id"],intro=movie_info[i]["intro"])
    index.index_doc(doc)

search_term = input("Enter term to search: ")
result = index.lookup_query(search_term)
rank={}
true=[]
prediction=[]

```

Enter term to search: 葉問
共符合12 indexing:7878

1230 (0.00022081520320408666):具有大規模毀滅能力，來自異次元的巨大怪獸，以及人類為了消滅牠們而打造，由人類駕駛的超巨型機器人之間曾經展開一場在全
915 (0.00020540782472984668):★《殺破狼》《葉問》製作團隊最新力作★挑戰地表最強老爸！古天樂首部動作武打片★動作場面全面升級殺戮是唯一的救贖睽違十
2747 (0.00018548536929781554):★《葉問》系列甄子丹及黃百鳴監製，電影武術大師袁和平執導，武術生張晉突破極限，挑戰完美功夫身手★2018釜山影展閉幕前
4553 (0.00017948150439354568):★聖誕跨年最強IP，葉問十週年精彩完結篇★甄子丹宗師回歸，十年傳奇最後一戰★《葉問》原班人馬再創巔峰，葉問為拯救李小冉
4869 (0.0001745151536592866):★《葉問1》《葉問2》正宗演員回歸前傳！★在葉問成為一代宗師之前，不為人知的故事…★杜宇航近身肉搏，再次挑戰拳拳實實
168 (0.00016237797134365016):★5億打造、超強卡司，暑期壓軸強檔電影★開創華語魔幻電影新高度，打造非凡視覺效果★華語影壇超強卡司共襄盛舉，李連杰、甄
2510 (0.00015793031887932982):★迪士尼經典動畫《花木蘭》真人版登上大銀幕★仙女姐姐劉亦菲化身花木蘭★穿上戰袍挑戰高難度武打動作★集結亞洲巨星功夫
5704 (0.00014621034257445361):★《紅翼行動》《2槍斃命》製片打造全新動作鉅獻★《葉問4》美國武打巨星史考特艾金斯、《狙擊生死線》人氣男星雷恩菲利浦火
1578 (0.00013779778997496527):★集合火爆動作《Mrs. K》、挑戰道德尺度《以青春的名義》、懸疑刺激《搶紅》、溫暖勵志《決戰食神》、愛情喜劇《天生不對》
47 (0.0001110959638972549):★集結經典賀歲《家有喜事》《哪一天我們會飛》《惡人谷》《北京愛情故事》《我來自紐約》《愛在深秋》《王家欣》《開飯啦》
227 (5.98315813150519e-05):★韓流天王李敏鎬首度演出華語作品，展現動作身手，不忘賣萌，帥氣擔當★李敏鎬與鐘漢良韓港雙帥帥帥搶金賞，成為今年最有新鮮
3768 (4.3477162623502925e-05):★聖誕跨年最強IP，葉問十週年精彩完結篇★甄子丹宗師回歸，十年傳奇最後一戰★《葉問》原班人馬再創巔峰，葉問為拯救李小冉

```
Precision: 1.0
Recall: 1.0
```