

```

import jieba
import json

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

import random
movie_links_list=[]
movie_link=[]
for i in range(len(movie_info)):
    movie_link.append(movie_info[i]["movie_url"])
for i in range(len(movie_info)):
    links=[]
    for j in range(5):
        x = random.randint(6123,14000)
        url="https://movies.yahoo.com.tw/movieinfo_main/"+str(x)
        if url in movie_link:
            links.append(url)
        else:
            j-=1
    movie_links_list.append(links)

connect={}
for i in range(len(movie_link)):
    connect.update({movie_link[i]:movie_links_list[i]})

import numpy as np
import networkx as nx

def nodes2matrix(node_json):
    node2id = {}
    dim = len(node_json)
    for id_ , key in enumerate(node_json.keys()):
        node2id[key] = id_
    matrix = np.zeros((dim,dim))
    for key in node_json.keys():
        nodeid = node2id[key]
        for neighbor in node_json[key]:
            neighborid = node2id[neighbor]
            matrix[neighborid][nodeid] = 1
    for i in range(dim):
        matrix[:,i] = matrix[:,i]/sum(matrix[:,i])
    return matrix

def pagerank(matrix, iter_ = 10, d=0.85):
    length = len(matrix[0])
    inital_value = np.ones(length)/length
    pagerank_value = inital_value
    for i in range(10):
        pagerank_value = matrix@pagerank_value* d + (1-d)/length
    return pagerank_value

matrix = nodes2matrix(connect)

my_rank=pagerank(matrix)

movie_all_info=[]
for i in range(len(movie_info)):
    id=i+1
    ch=movie_info[i]["ch_name"]
    en=movie_info[i]["en_name"]
    url=movie_info[i]["movie_url"]
    date=movie_info[i]["release_date"]
    intro=movie_info[i]["intro"]
    links=movie_links_list[i]
    my_pagerank=my_rank[i]

```

```
label=movie_info[i]["label"]
movie_all_info.append({"id":id,"ch_name":ch,"en_name":en,"url":url,"release_date":date,"intro":intro,"links":links,"label":label,"pagera
with open("movie_all_info.json","w") as f:
    json.dump(movie_all_info,f)

print(movie_all_info[0])

vieinfo_main/6635', 'https://movies.yahoo.com.tw/movieinfo_main/9361'], 'label': ['劇情', '犯罪'], 'pagerank': 0.00015658648769140712}
```

Colab 付費產品 - [按這裡取消合約](#)