

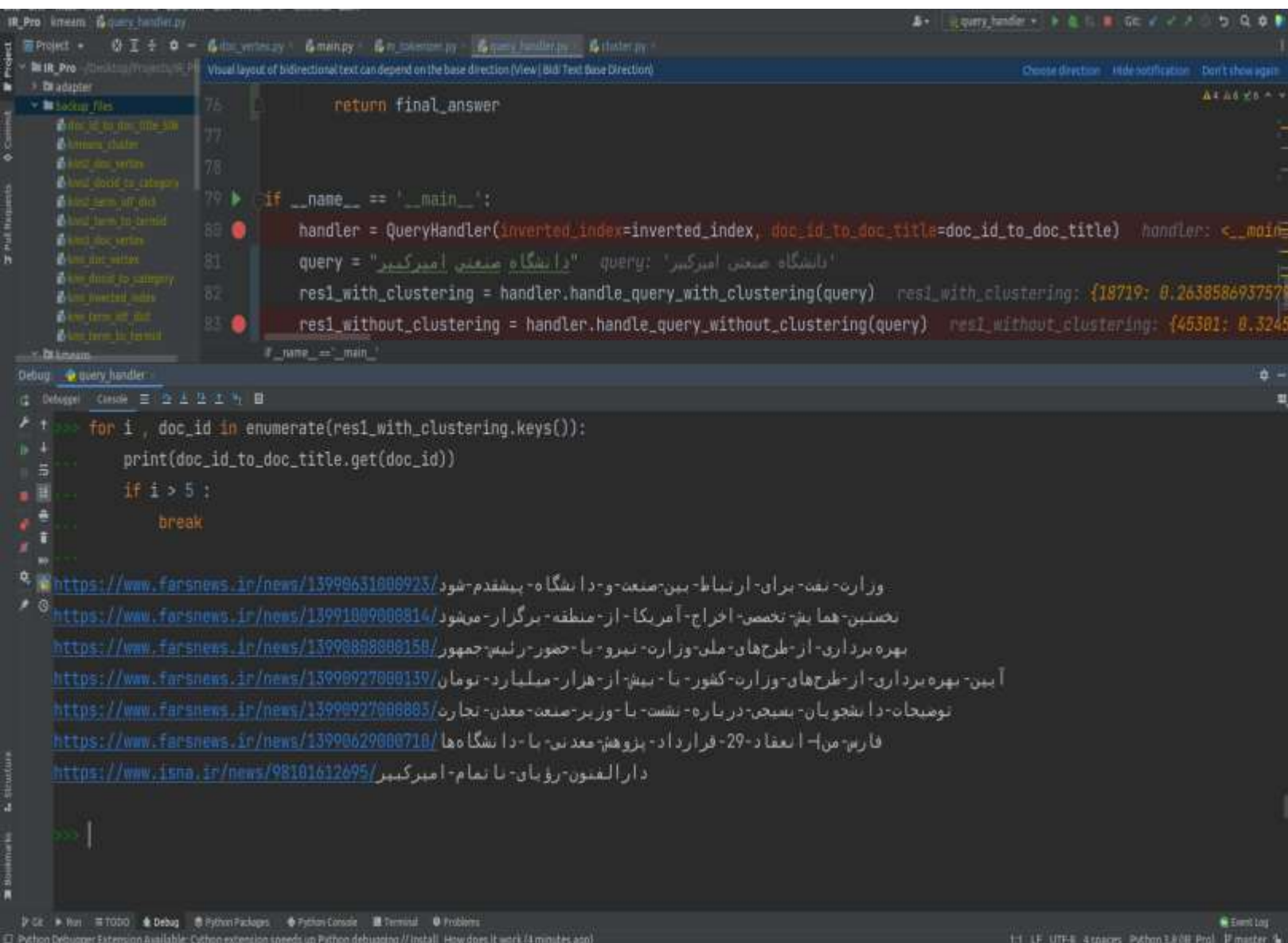
[Date]

# گزارش فاز سوم بازیابی اطلاعات

امیرحسن سعادت‌مند 9731029

با خوشه بندی (clustering)

پرسمان : دانشگاه صنعتی امیرکبیر

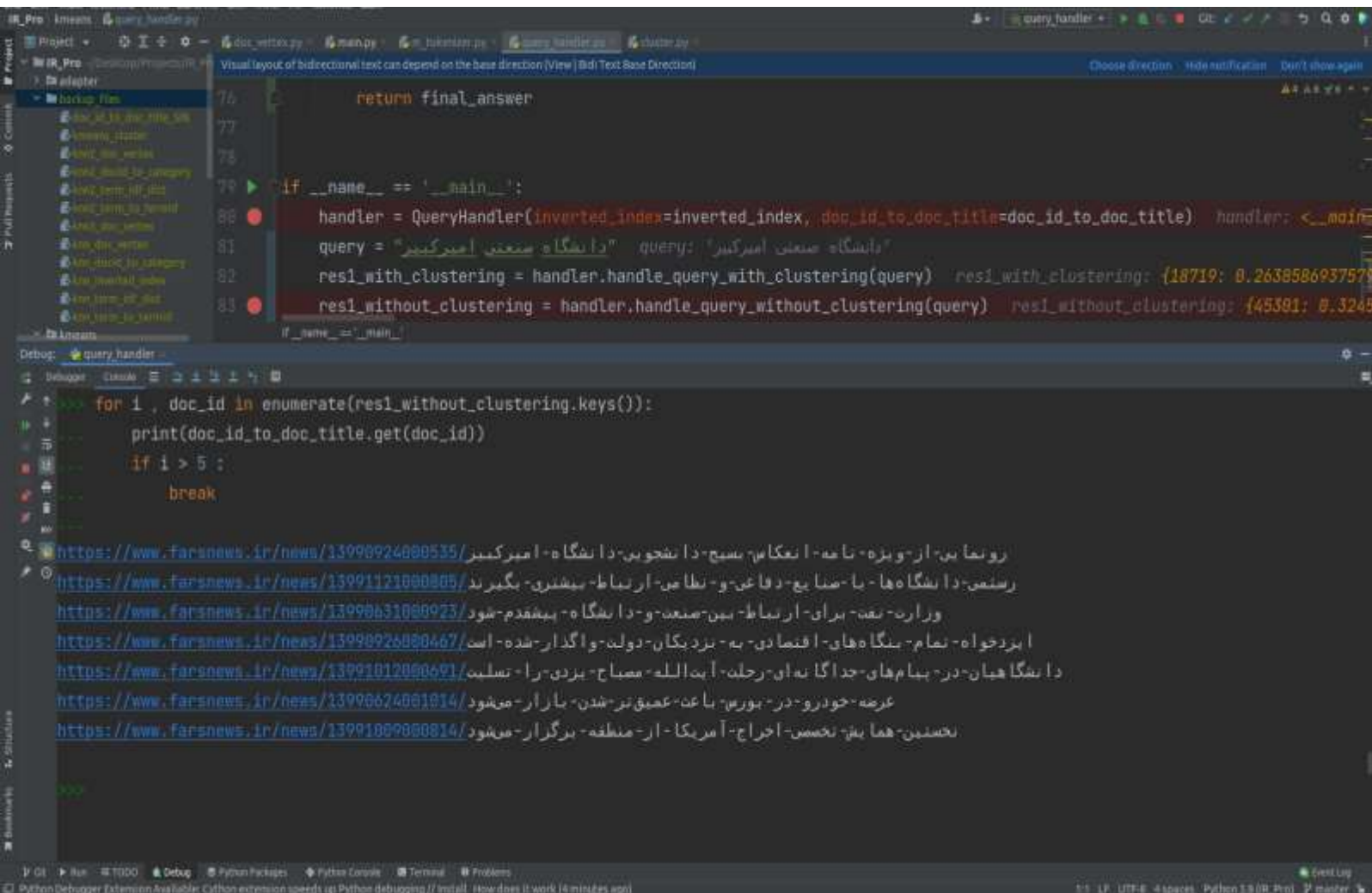


```
76     return final_answer
77
78
79 if __name__ == '__main__':
80     handler = QueryHandler(inverted_index=inverted_index, doc_id_to_doc_title=doc_id_to_doc_title) handler: <_main__
81     query = "دانشگاه صنعتی امیرکبیر" query: 'دانشگاه صنعتی امیرکبیر'
82     res1_with_clustering = handler.handle_query_with_clustering(query) res1_with_clustering: {18719: 0.2638586937578
83     res1_without_clustering = handler.handle_query_without_clustering(query) res1_without_clustering: {45581: 0.3245

if __name__ == '__main__':

Debug: query_handler
for i, doc_id in enumerate(res1_with_clustering.keys()):
    print(doc_id_to_doc_title.get(doc_id))
    if i > 5:
        break
https://www.farsnews.ir/news/13990631000923/وزارت نفت برای ارتباط بین صنعت و دانشگاه پیشقدم شود
https://www.farsnews.ir/news/13991009000814/نخستین هماهنگی تخصصی اخراج آمریکا از منطقه برگزار میشود
https://www.farsnews.ir/news/13990808000158/بهره برداری از طرح های ملی وزارت نیرو با حضور رئیس جمهور
https://www.farsnews.ir/news/13990927000139/آیین بهره برداری از طرح های وزارت کشور تا پیش از هزار میلیارد تومان
https://www.farsnews.ir/news/13990927000803/توضیحات دانشجویان بسیجی درباره نشست با وزیر صنعت معدن تجارت
https://www.farsnews.ir/news/13990629000710/فارس من ایام عباد 29 قرارداد پژوهش معدنی با دانشگاه ها
https://www.isna.ir/news/98101612695/دارالفنون رؤیای نا تمام امیرکبیر
```

بدون خوشه بندی :



The screenshot shows a Python IDE with a project named 'IR\_Pre'. The main editor displays the 'query\_handler.py' file. The code defines a 'QueryHandler' class and a main function. The main function initializes the handler with an inverted index and doc\_id\_to\_doc\_title mapping, then processes a query in Persian. It uses 'handle\_query\_with\_clustering' and 'handle\_query\_without\_clustering' methods. The debug console shows the output of the 'handle\_query\_without\_clustering' method, which prints a list of document IDs and their corresponding titles from the 'doc\_id\_to\_doc\_title' mapping.

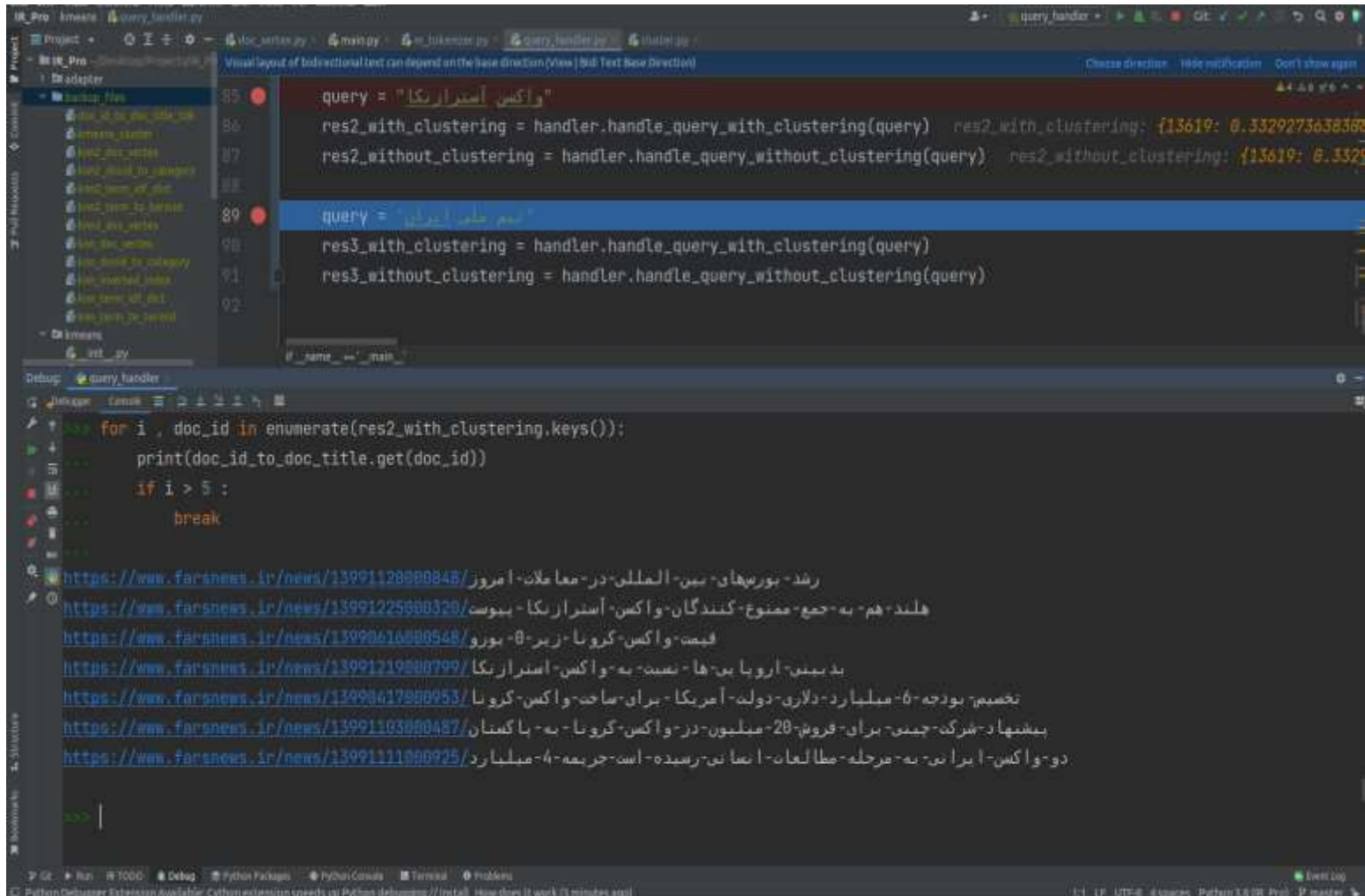
```
76     return final_answer
77
78
79 if __name__ == '__main__':
80     handler = QueryHandler(inverted_index=inverted_index, doc_id_to_doc_title=doc_id_to_doc_title)
81     query = "دانشگاه صنعتی امیرکبیر"
82     res1_with_clustering = handler.handle_query_with_clustering(query)
83     res1_without_clustering = handler.handle_query_without_clustering(query)
```

```
for i, doc_id in enumerate(res1_without_clustering.keys()):
    print(doc_id_to_doc_title.get(doc_id))
    if i > 5:
        break
```

https://www.farsnews.ir/news/13990924000535/امیرکبیر-دانشگاه-دانشجوی-دانشگاه-امیرکبیر  
https://www.farsnews.ir/news/13991121000805/رئیس-دانشگاه-با-صنایع-دفاعی-و-نظامی-ارتباط-بشود-بگیرند  
https://www.farsnews.ir/news/13990631000923/وزارت-نفت-برای-ارتباط-بین-صنعت-و-دانشگاه-پیشقدم-شود  
https://www.farsnews.ir/news/13990924000467/ایزدخواه-تمام-بنگاه-های-اقتصادی-به-بزرگان-دولت-واگذار-شده-است  
https://www.farsnews.ir/news/13991012000691/دانشگاه-هیان-در-پیام-های-جداگانه-ای-رجل-آیت-الله-مصباح-یزدی-را-تسلیم  
https://www.farsnews.ir/news/13990624001014/عرشه-خودرو-در-بورس-با-عن-عمیق-تر-شدن-بازار-میشود  
https://www.farsnews.ir/news/13991009000814/نخستین-هماهنگی-نخستین-اجرا-آمریکا-از-منطقه-برگزار-میشود

پرسمان : واكسن آسترازنكا

با خوشه بندى :



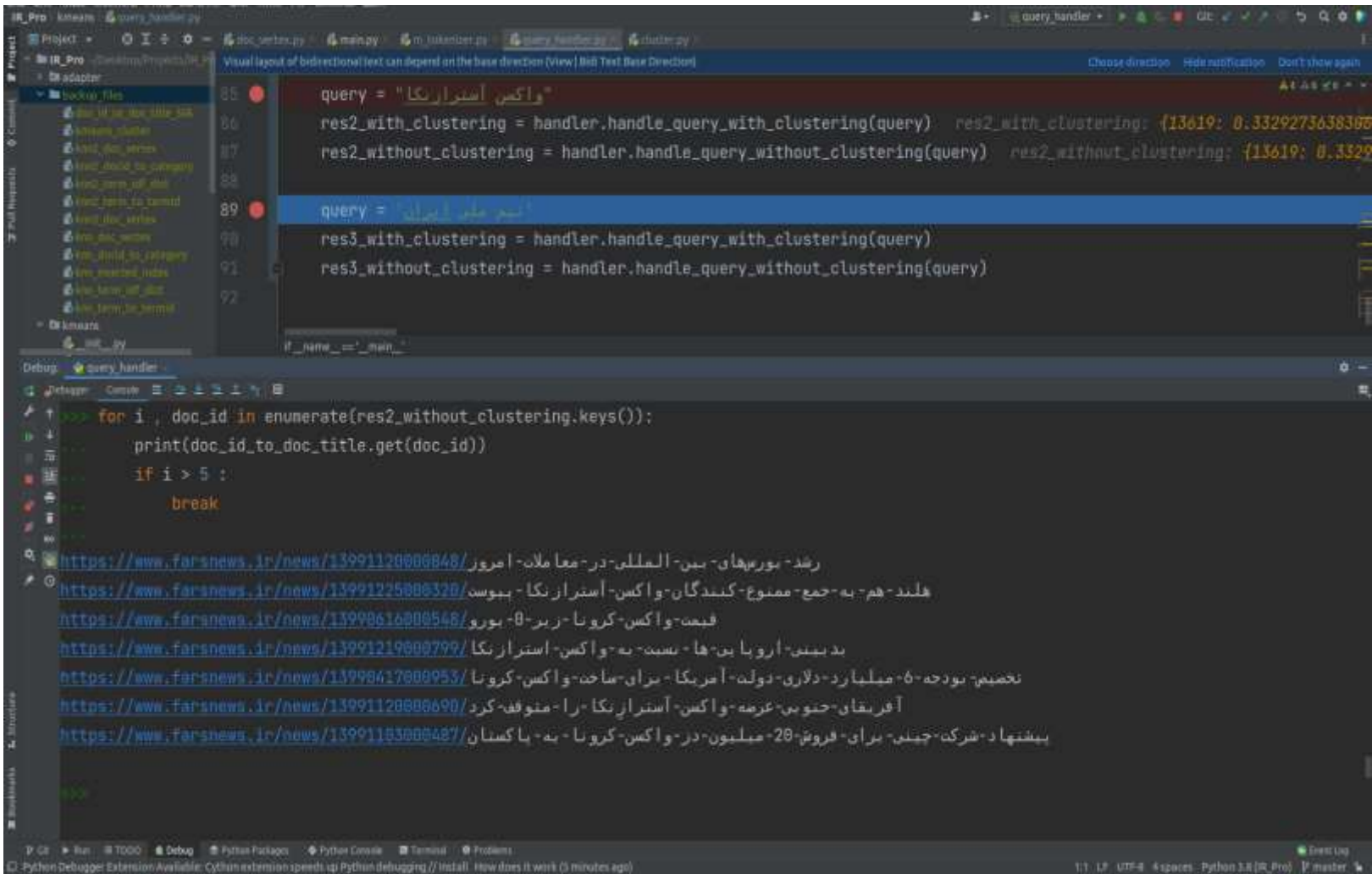
The screenshot shows a Python IDE with a project named 'IR\_Pro'. The main script, 'query\_handler.py', is open and shows the following code:

```
85 query = "واكسن آسترازنكا"
86 res2_with_clustering = handler.handle_query_with_clustering(query)
87 res2_without_clustering = handler.handle_query_without_clustering(query)
88
89 query = "نعم طاهر ايرانى"
90 res3_with_clustering = handler.handle_query_with_clustering(query)
91 res3_without_clustering = handler.handle_query_without_clustering(query)
92
```

The debug console shows the output of the script, displaying a list of news URLs and their titles, along with a list of news titles and their URLs. The output is as follows:

```
for i, doc_id in enumerate(res2_with_clustering.keys()):
    print(doc_id_to_doc_title.get(doc_id))
    if i > 5:
        break
https://www.farsnews.ir/news/13991128888848/ رشد- بورسهای- بین- المللی- در- معاملات- امروز
https://www.farsnews.ir/news/13991225688328/ هلند- هم- به- جمع- متنوع- کنندگان- واكسن- آسترازنكا- پیوست
https://www.farsnews.ir/news/13998616888948/ قیمت- واكسن- کرونا- زیر- 8- یورو
https://www.farsnews.ir/news/13991219888799/ بدبینی- اروپا- بی- نها- نسبت- به- واكسن- آسترازنكا
https://www.farsnews.ir/news/13998617888953/ تخمین- بودجه- 6- میلیارد- دلاری- دولت- آمریکا- برای- ساخت- واكسن- کرونا
https://www.farsnews.ir/news/139911038880487/ بیشنهاد- شرکت- چینی- برای- فروش- 28- میلیون- در- واكسن- کرونا- به- پاکستان
https://www.farsnews.ir/news/13991111888925/ دو- واكسن- ا- برا- نی- به- مرحله- مطالعات- ا- نما- نی- رسیده- است- جریمه- 4- میلیارد
```

بدون خوشه بندی :



```
Project: IR_Pro - C:\Users\Project\IR_Pro
doc_corpus.py | main.py | m_tokenizer.py | query_handler.py | cluster.py
Visual layout of bidirectional text can depend on the base direction (view | Bid Text Base Direction)
Choose direction | Hide notification | Don't show again

85 query = "واکس استرازا"
86 res2_with_clustering = handler.handle_query_with_clustering(query) res2_with_clustering: {13619: 0.3329273638367
87 res2_without_clustering = handler.handle_query_without_clustering(query) res2_without_clustering: {13619: 0.3329
88
89 query = "قیمت دلار ایران"
90 res3_with_clustering = handler.handle_query_with_clustering(query)
91 res3_without_clustering = handler.handle_query_without_clustering(query)
92

if __name__ == '__main__':

Debug: query_handler.py
>>> for i, doc_id in enumerate(res2_without_clustering.keys()):
print(doc_id_to_doc_title.get(doc_id))
if i > 5 :
break

https://www.farsnews.ir/news/13991120006848/ رشد-بورسهای-بین-المللی-در-معاملات-امروز
https://www.farsnews.ir/news/13991225000320/ هلند-هم-به-جمع-ممنوع-کنندگان-واکس-استرازا-نکا-ببوست
https://www.farsnews.ir/news/13990616000548/ قیمت-واکس-کرونا-زیر-8-یورو
https://www.farsnews.ir/news/13991219000799/ بدبینی-اروپا-بن-ها-نسبت-به-واکس-استرازا-نکا
https://www.farsnews.ir/news/13990417000953/ تصمیم-بودجه-6-میلیارد-دلاری-دولت-آمریکا-برای-ساخت-واکس-کرونا
https://www.farsnews.ir/news/13991120006690/ آفریقای-جنوبی-عزمه-واکس-استرازا-نکا-را-متوقف-کرد
https://www.farsnews.ir/news/13991163000487/ پیشنهاد-شرکت-چینی-برای-فروش-20-میلیون-در-واکس-کرونا-به-پاکستان

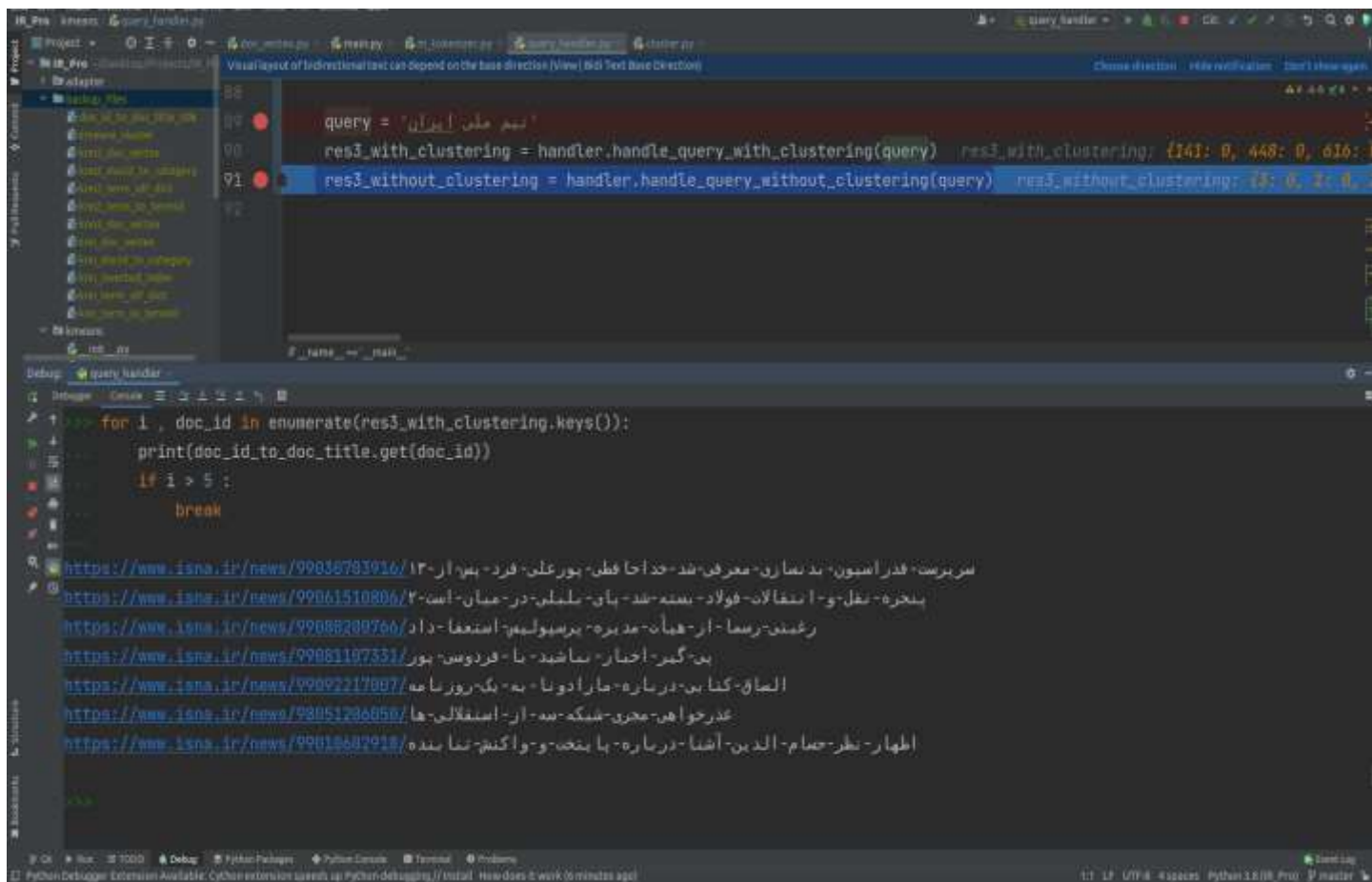
>>>
```

Python Debugger Extension Available: CPython extension speeds up Python debugging // Install. How does it work (3 minutes ago)

1:1 LF UTF-8 4 spaces Python 3.8 (R\_Pro) master



## پرسمان : تیم ملی ایران با خوشه بندی :



```
query = "تیم ملی ایران"
res3_with_clustering = handler.handle_query_with_clustering(query)
res3_without_clustering = handler.handle_query_without_clustering(query)
```

```
for i, doc_id in enumerate(res3_with_clustering.keys()):
    print(doc_id_to_doc_title.get(doc_id))
    if i > 5:
        break
```

<https://www.isna.ir/news/99036783916/۱۲-بهر-ار-سرپرست-فدراسیون-بدنسازی-معرفی-شد-خداحافظی-بورعلی-فرد-بهر-ار>

<https://www.isna.ir/news/99061510806/۲-پنجره-نقل-و-انتقالان-فولاد-بسته-شد-بازی-لیگ-در-میان-است-۲>

<https://www.isna.ir/news/99088289766/رغمین-رسم-ار-همان-مدیر-پرسونل-استعفا-داد>

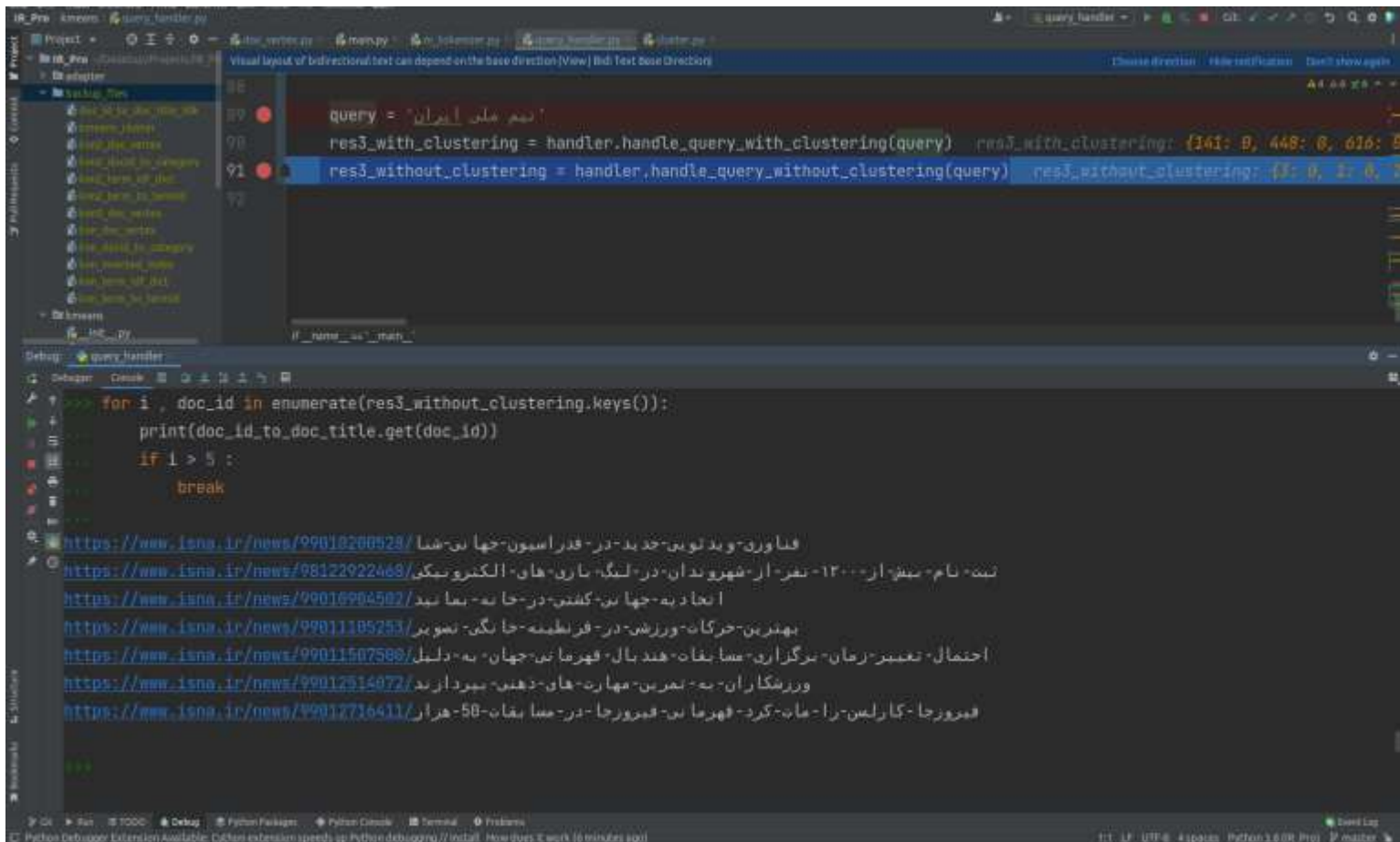
<https://www.isna.ir/news/99061107331/بهر-گیر-اخبار-ماشهد-با-فردوس-بور>

<https://www.isna.ir/news/99092217007/الماق-کنا-بهر-در-باز-مارادونا-به-بهر-روز-ما-ه>

<https://www.isna.ir/news/99051206050/عذرخواهی-محر-شیکه-به-ار-استغالی-ها>

<https://www.isna.ir/news/99016682918/اطهار-نظر-حام-الدین-اشنا-در-باز-با-بهر-و-واکنش-نما-بهر>

## بدون خوشه بندی :



```
query = "نم علی ایران"
res3_with_clustering = handler.handle_query_with_clustering(query)
res3_without_clustering = handler.handle_query_without_clustering(query)
```

```
for i, doc_id in enumerate(res3_without_clustering.keys()):
    print(doc_id_to_doc_title.get(doc_id))
    if i > 5:
        break
```

<https://www.isna.ir/news/99010200928/> فناوری-ویدئویی-جدید-در-قدراسیون-چها-بی-شما

<https://www.isna.ir/news/98122922468/> ثبت-نام-بیش-از-1۴۰۰-تغیر-از-فهرست-ان-در-لیگ-بارق-های-الکترونیکی

<https://www.isna.ir/news/99010904502/> اتحادیه-چها-بی-گشتی-در-خا-به-بما-نید

<https://www.isna.ir/news/99011105253/> بهترین-حرکات-ورزشی-در-فرهنگ-خا-نگی-نمویز

<https://www.isna.ir/news/99011507586/> احتمال-تعبیر-زمان-برگزاری-مسابقه-هندبال-فهرما-بی-چها-به-دلیل

<https://www.isna.ir/news/99012514072/> ورزشکاران-به-تمرین-مهارت-های-ذهنی-بپردازند

<https://www.isna.ir/news/99012716411/> فیروزجا-کارلسن-را-مات-کرد-فهرما-بی-فیروزجا-در-مسابقه-5۵-هزار

در خوشه بندی تعداد دسته ها را ( $k = 5$ ) در نظر گرفتیم و  $b = 2$  که در نتیجه داده های موجود در 2 خوشه را با تابع شباهت بررسی میکند

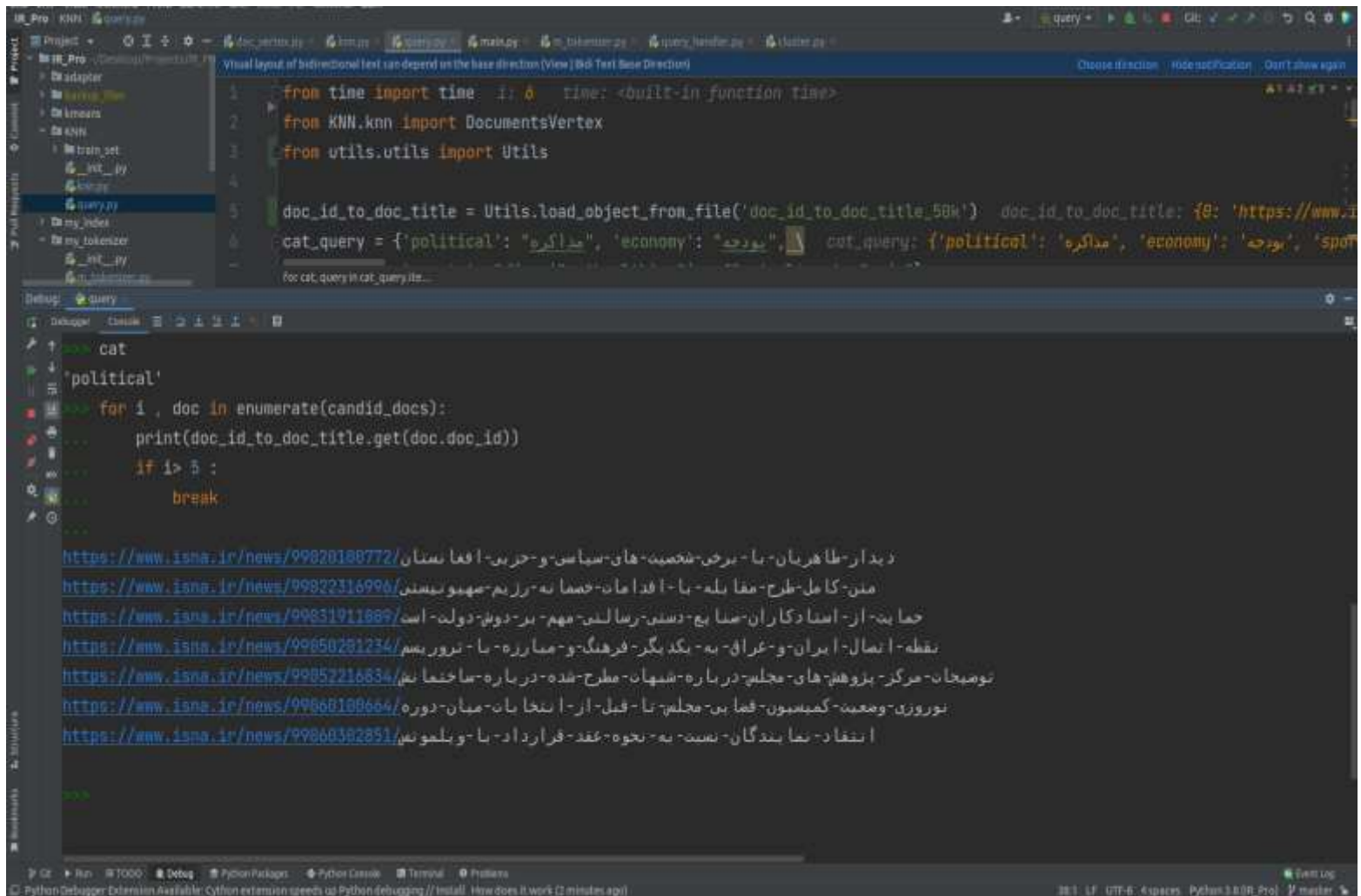
مشخص است با خوشه بندی و بدون آن نتیجه آنچنان متفاوت نیست اما سرعت اجرای پرسمان بسیار کاهش یافته است

به دلیل آنکه فقط 2/5 داده ها که در خوشه این دو مرکز هستند بررسی میشوند

نکته دیگر آنست که در خوشه بندی 5 بار reassignment انجام شده و داد های موجود در خوشه با توزیع تقریباً یکسانی هستند ( به لحاظ تعداد )

With category = political

Query = مذاکره



```
from time import time; 0 time: <built-in function time>
from KNN.knn import DocumentsVertex
from utils.utils import Utils

doc_id_to_doc_title = Utils.load_object_from_file('doc_id_to_doc_title_50k') doc_id_to_doc_title: {0: 'https://www.1
cat_query = {'political': 'مذاکره', 'economy': 'بازچه', 'sport': 'مذاکره', 'economy': 'بازچه', 'spot

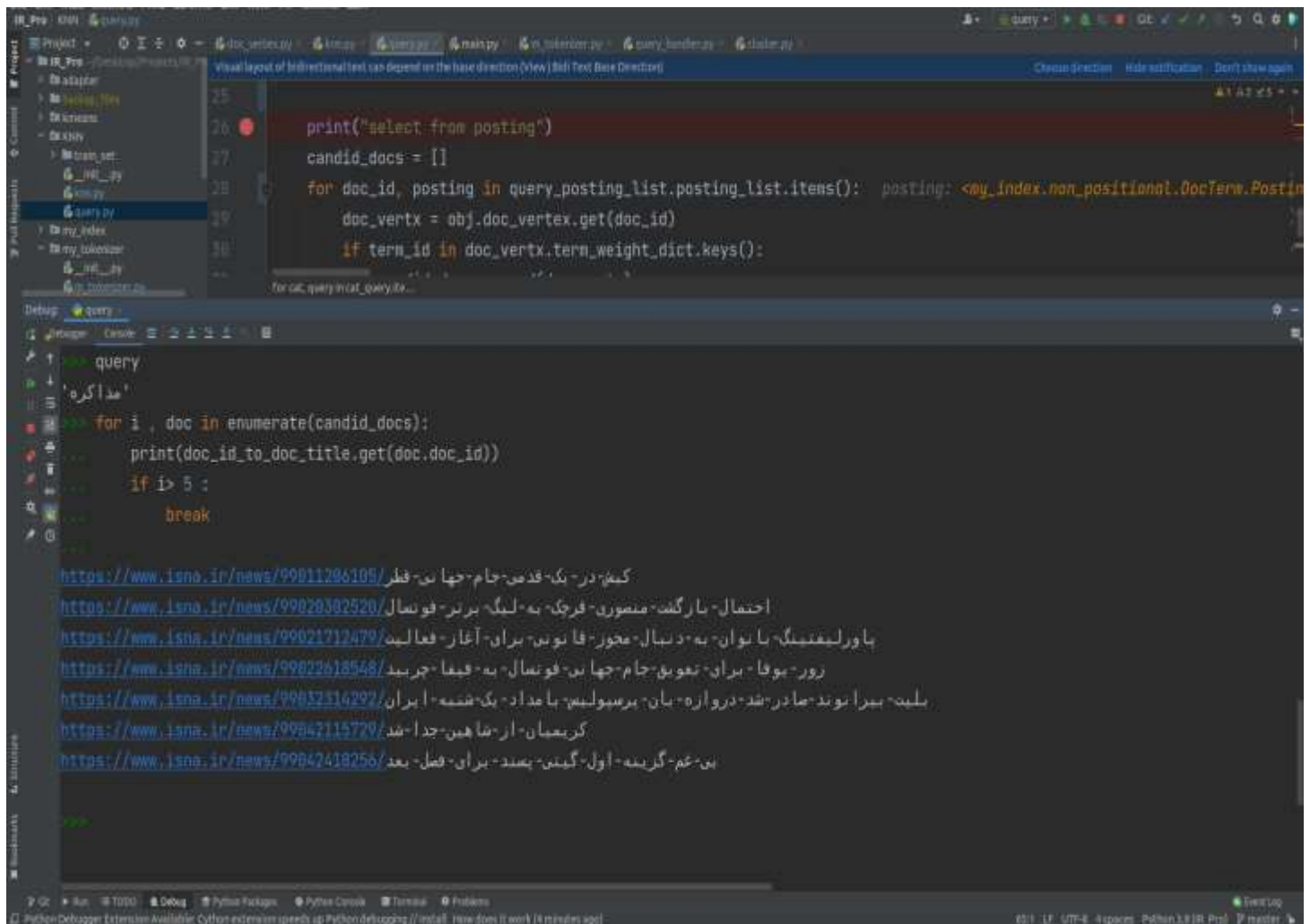
for cat, query in cat_query.items():
    for i, doc in enumerate(candid_docs):
        print(doc_id_to_doc_title.get(doc.doc_id))
        if i > 5:
            break

https://www.isna.ir/news/99820188772/ دیدار طاهریان-با-یروح-شعبیه-های-سیاسی-و-حزب-افغانستان
https://www.isna.ir/news/99822316996/ ملن-کامل-طرح-مقابله-با-افغان-خسما-نه-رزیم-مهیوینسین
https://www.isna.ir/news/99831911889/ حما-پن-از-اسنادکاران-صایع-دستی-رساله-مهم-بر-دوش-دولت-است
https://www.isna.ir/news/99850281234/ نقطه-انمال-ایران-و-عراق-به-یکدیگر-فرهنگ-و-مبارزه-با-تروریسم
https://www.isna.ir/news/99852216834/ توضیحات-مرکز-پژوهش-های-مجلس-درباره-شبهات- مطرح-شده-درباره-ساخته-نم
https://www.isna.ir/news/99860108664/ روزی-وضعیت-کمیسیون-قما-بی-مجلس-تا-قبل-از-انتخابان-میان-دوره
https://www.isna.ir/news/99860382851/ انتقاد-نمایندگان-نست-به-نحوه-عقد-قرارداد-با-ویلمون
```



Without category

Query = مذاکره

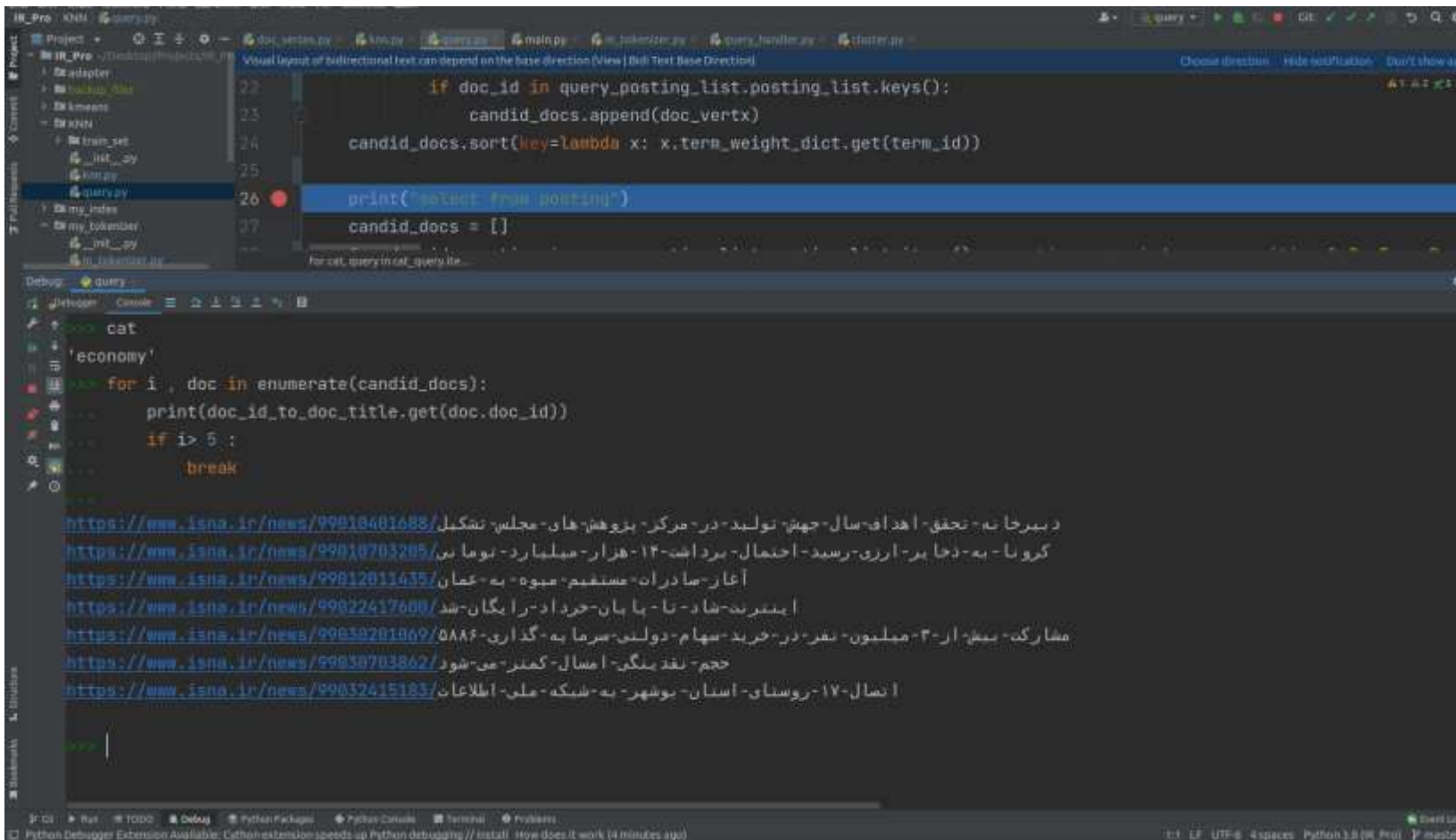


```
25
26 print("select from posting")
27 candid_docs = []
28 for doc_id, posting in query_posting_list.posting_list.items():
29     doc_vertx = obj.doc_vertx.get(doc_id)
30     if term_id in doc_vertx.term_weight_dict.keys():
31         for cat, query in cat_query_dict.items():
```

```
>>> query
'مذاکره'
>>> for i, doc in enumerate(candid_docs):
...     print(doc_id_to_doc_title.get(doc.doc_id))
...     if i > 5:
...         break
...
https://www.isna.ir/news/99011206105/کمپین-در-یک-قدح-جام-چها-نی-فطر
https://www.isna.ir/news/99020302520/احتمال-بارگشت-منصور-فرچکه-به-لیگ-برتر-فوتسال
https://www.isna.ir/news/99021712479/پارلیفینگ-با-توان-به-دنیال-مخور-فا-تونی-برای-آغاز-فعالیت
https://www.isna.ir/news/99022618548/زور-یوفا-برای-تعویق-جام-چها-نی-فوتسال-به-فیفا-جرید
https://www.isna.ir/news/99032314292/بلیت-بیرا-نوند-ما-در-شد-دروازه-بان-برسیولیس-با-مداد-یک-شنبه-ا-بران
https://www.isna.ir/news/99042115729/کریمیان-از-شا-هین-جدا-شد
https://www.isna.ir/news/99042418256/بی-غم-گریبه-اول-گیتی-بسد-برای-فضل-بعد
```

Category = economy

Query = بودجه



The screenshot shows a Python IDE with a project named 'KHAN'. The file explorer on the left lists several files, including 'query.py'. The main editor window displays the following Python code:

```
22 if doc_id in query_posting_list.posting_list.keys():
23     candid_docs.append(doc_vctx)
24     candid_docs.sort(key=lambda x: x.term_weight_dict.get(term_id))
25
26 print("select from posting")
27 candid_docs = []
```

Below the code editor, the 'Debug' console shows the execution of the script. The output is as follows:

```
cat
'economy'
for i, doc in enumerate(candid_docs):
    print(doc_id_to_doc_title.get(doc.doc_id))
    if i > 5:
        break
```

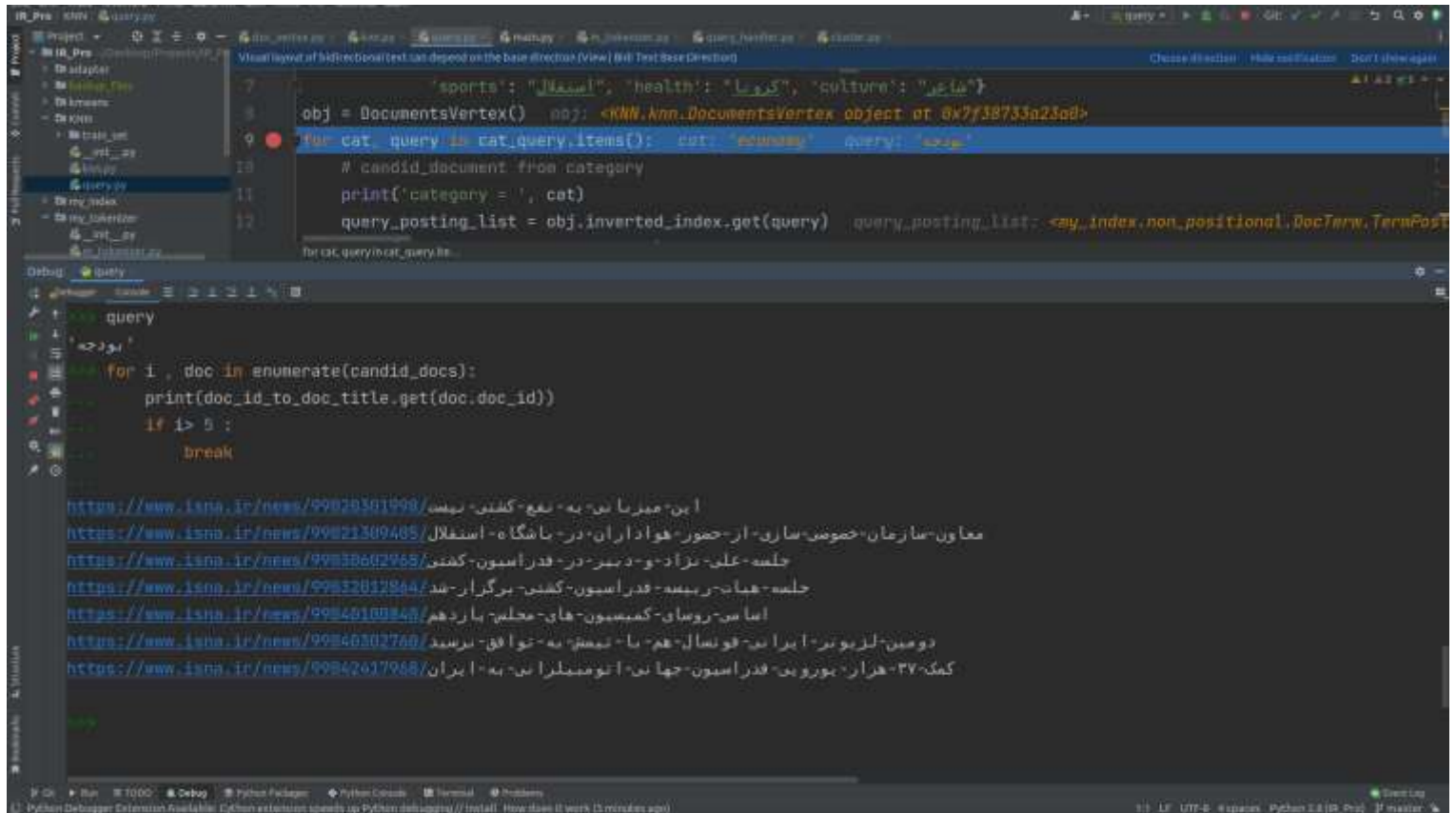
The output lists several news URLs from the 'isna.ir' domain, all related to the 'economy' category. The URLs are:

- <https://www.isna.ir/news/99010401688/>
- <https://www.isna.ir/news/99010703205/>
- <https://www.isna.ir/news/99012011435/>
- <https://www.isna.ir/news/99022417600/>
- <https://www.isna.ir/news/99030201069/>
- <https://www.isna.ir/news/99030703862/>
- <https://www.isna.ir/news/99032415183/>

The IDE interface includes a 'Run' button, a 'Debug' button, and a 'Python Debugger Extension Available' notification at the bottom.

Without category

Query = بودجه



```
7 'sports': "اسپورٹس", 'health': "کھانا", 'culture': "سفر"}
8 obj = DocumentsVertex() obj: <KNN.knn.DocumentsVertex object at 0x7f38733a230d>
9 for cat, query in cat_query.items(): cat: 'economy' query: 'بودجه'
10 # candid_document from category
11 print('category = ', cat)
12 query_posting_list = obj.inverted_index.get(query) query_posting_list: <my_index.non_positional.DocTerm.TermPost>

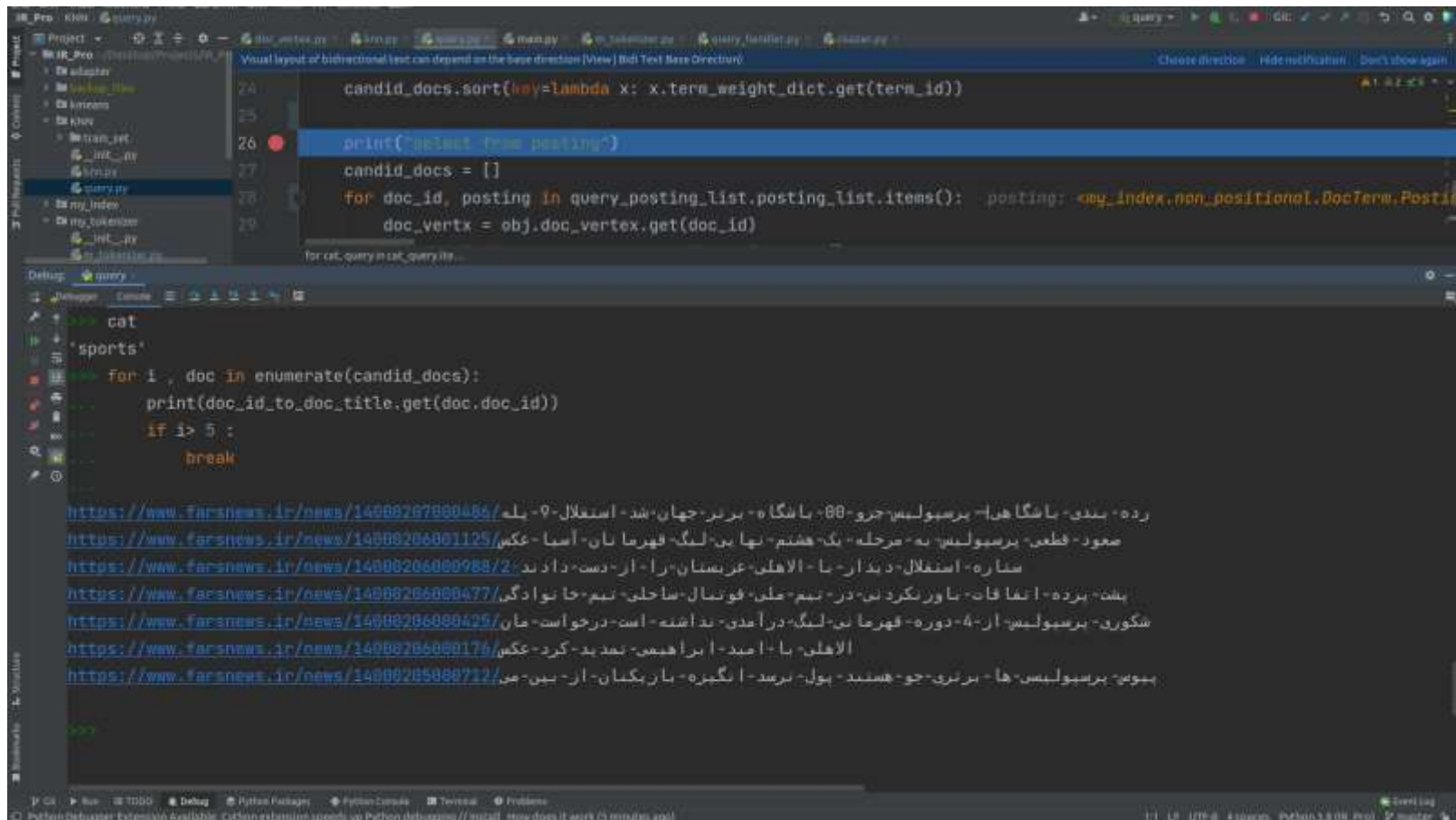
for cat, query in cat_query.items():

Debug: query
query
بودجه
for i, doc in enumerate(candid_docs):
    print(doc_id_to_doc_title.get(doc.doc_id))
    if i > 5:
        break

https://www.isna.ir/news/99020301998/این-میراث-بی-بی-نوع-کشتی-نیست
https://www.isna.ir/news/99021309405/معاون-سازمان-خصوصی-سازی-از-حمور-خواداران-در-باشگاه-استقلال
https://www.isna.ir/news/99030602965/جلسه-علی-نژاد-و-دبیر-در-فدراسیون-کشتی
https://www.isna.ir/news/99032012864/جلسه-عباس-رییسی-فدراسیون-کشتی-برگزار-شد
https://www.isna.ir/news/99040100840/اساس-روسای-کمیسیون-های-مجلس-باردهم
https://www.isna.ir/news/99040302760/دومین-لژیونر-ایران-فوتسال-هم-با-تیمش-به-توافقی-ترسید
https://www.isna.ir/news/99042617968/کمک-۳۷-هزار-یورویی-فدراسیون-جها-نی-آتمیلرا-بی-به-ایران
```

Category = Sport

Query = استقلال



The screenshot shows a Python IDE with a project named 'IR\_Pro'. The main editor displays a script for processing a query. The script sorts candidate documents by term weight, prints a debug message, and iterates through a list of postings to retrieve document vertices. The output console shows the results for the query 'استقلال' (Esteghlal), listing document IDs and their corresponding titles in Persian.

```
24 candid_docs.sort(key=lambda x: x.term_weight_dict.get(term_id))
25
26 print("select from posting")
27 candid_docs = []
28 for doc_id, posting in query_posting_list.posting_list.items():
29     doc_vertx = obj.doc_vertex.get(doc_id)
```

Debug: query

```
>>> cat
'sports'
>>> for i, doc in enumerate(candid_docs):
...     print(doc_id_to_doc_title.get(doc.doc_id))
...     if i > 5:
...         break
```

https://www.farsnews.ir/news/14000267800486/ رده بندی باشگاهها- برسیولیس-جرو-80- باشگاه- برتری-جهان-شد- استقلال-9- یله  
https://www.farsnews.ir/news/14000266001125/ صعود- قطعی- برسیولیس- به- مرحله- یک- هشتم- بها- بن- لنگ- قهرمان- آسیا- عکس  
https://www.farsnews.ir/news/14000266000988/2/ ستاره- استقلال- دیدار- با- الاهلی- عربستان- را- از- دست- داد- بند  
https://www.farsnews.ir/news/14000266000477/ پشت- برده- اتفا- قات- باور- نکرد- بن- در- نیم- علی- فوتبال- ساحلی- نیم- جا- نوا- دگی  
https://www.farsnews.ir/news/14000266000425/ شکوری- برسیولیس- از- 4- دوره- قهرمان- بن- لنگ- در- آمدی- بد- آینه- است- درخواست- مان  
https://www.farsnews.ir/news/14000266000176/ الاهلی- با- امید- ا- برا- هیمی- تمد- بد- کرد- عکس  
https://www.farsnews.ir/news/14000265000712/ پیوست- برسیولیس- ها- برتری- جو- هستند- پول- نرسند- ا- نگیزه- با- ریکان- از- بین- می

Query = استقلال

The screenshot displays a Jupyter Notebook environment with a Python script and its execution output. The script defines a query and iterates over candidate documents to find the top 5 results based on a specific query.

**Script Code:**

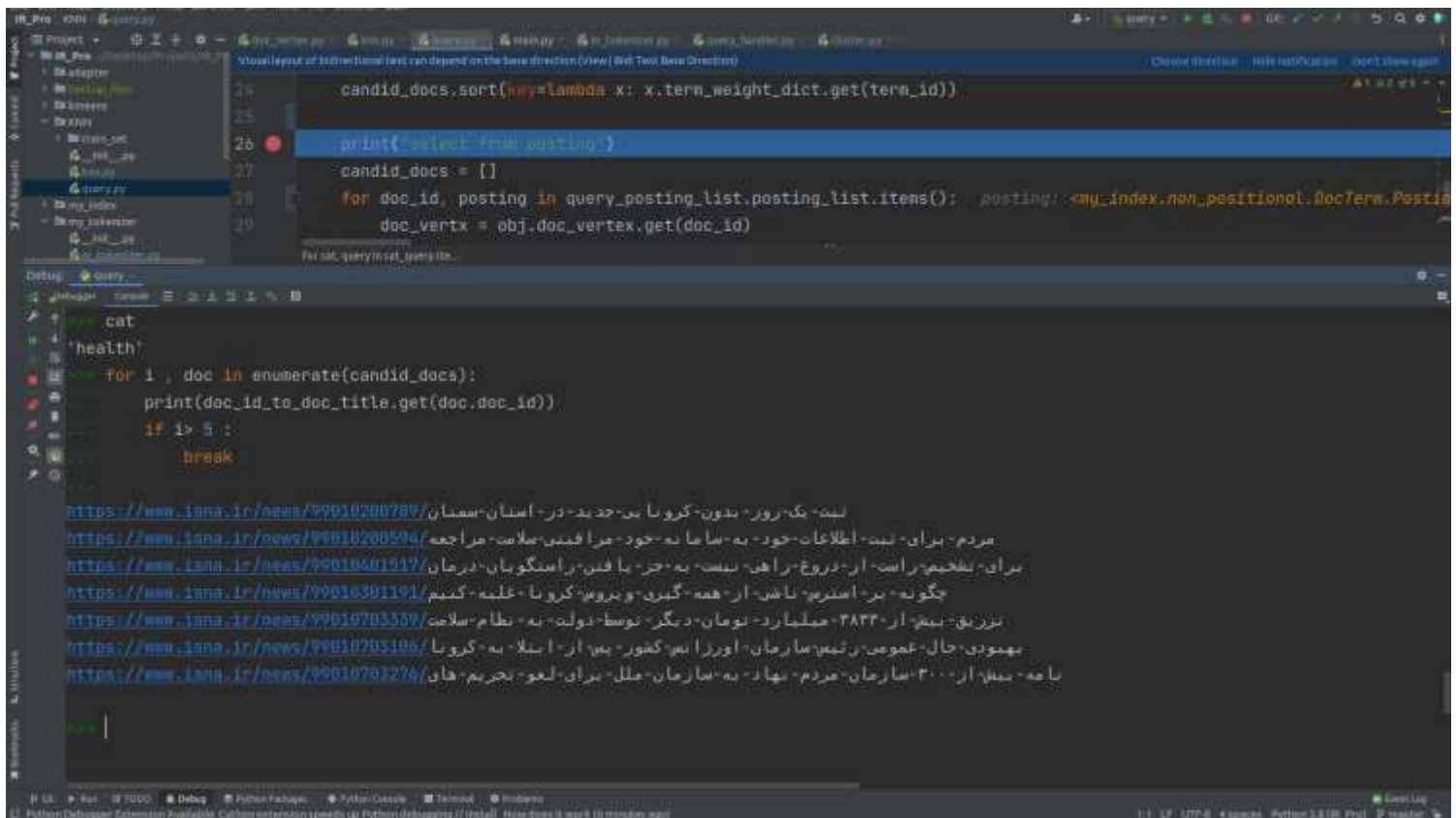
```

1 'sports': "استقلال", 'health': "کرونا", 'culture': "آغار"}
2 obj = DocumentsVertex() obj: <KNW.knn.DocumentsVertex object at 0x7f38735a2308>
3 for cat, query in cat_query.items(): cat: 'sports' query: "استقلال"
4 # candid document from category
5 print('category = ', cat)
6 query_posting_list = obj.inverted_index.get(query) query_posting_list: <my_index.non_positional.DocTerm.TermPost
7
8 for cat, query in cat_query.items():
9     query_posting_list = obj.inverted_index.get(query)
10     for doc_id, doc_title in enumerate(query_posting_list):
11         print(doc_id, doc_title)
12         if i > 5:
13             break
14
15 query
16 'استقلال'
17 for i, doc in enumerate(candid_docs):
18     print(doc_id_to_doc_title.get(doc.doc_id))
19     if i > 5:
20         break
21
22 query
23 'استقلال'
24 for i, doc in enumerate(candid_docs):
25     print(doc_id_to_doc_title.get(doc.doc_id))
26     if i > 5:
27         break
28
29 query
30 'استقلال'
31 for i, doc in enumerate(candid_docs):
32     print(doc_id_to_doc_title.get(doc.doc_id))
33     if i > 5:
34         break
35
36 query
37 'استقلال'
38 for i, doc in enumerate(candid_docs):
39     print(doc_id_to_doc_title.get(doc.doc_id))
40     if i > 5:
41         break
42
43 query
44 'استقلال'
45 for i, doc in enumerate(candid_docs):
46     print(doc_id_to_doc_title.get(doc.doc_id))
47     if i > 5:
48         break
49
50 query
51 'استقلال'
52 for i, doc in enumerate(candid_docs):
53     print(doc_id_to_doc_title.get(doc.doc_id))
54     if i > 5:
55         break
56
57 query
58 'استقلال'
59 for i, doc in enumerate(candid_docs):
60     print(doc_id_to_doc_title.get(doc.doc_id))
61     if i > 5:
62         break
63
64 query
65 'استقلال'
66 for i, doc in enumerate(candid_docs):
67     print(doc_id_to_doc_title.get(doc.doc_id))
68     if i > 5:
69         break
70
71 query
72 'استقلال'
73 for i, doc in enumerate(candid_docs):
74     print(doc_id_to_doc_title.get(doc.doc_id))
75     if i > 5:
76         break
77
78 query
79 'استقلال'
80 for i, doc in enumerate(candid_docs):
81     print(doc_id_to_doc_title.get(doc.doc_id))
82     if i > 5:
83         break
84
85 query
86 'استقلال'
87 for i, doc in enumerate(candid_docs):
88     print(doc_id_to_doc_title.get(doc.doc_id))
89     if i > 5:
90         break
91
92 query
93 'استقلال'
94 for i, doc in enumerate(candid_docs):
95     print(doc_id_to_doc_title.get(doc.doc_id))
96     if i > 5:
97         break
98
99 query
100 'استقلال'
101 for i, doc in enumerate(candid_docs):
102     print(doc_id_to_doc_title.get(doc.doc_id))
103     if i > 5:
104         break
105
106 query
107 'استقلال'
108 for i, doc in enumerate(candid_docs):
109     print(doc_id_to_doc_title.get(doc.doc_id))
110     if i > 5:
111         break
112
113 query
114 'استقلال'
115 for i, doc in enumerate(candid_docs):
116     print(doc_id_to_doc_title.get(doc.doc_id))
117     if i > 5:
118         break
119
120 query
121 'استقلال'
122 for i, doc in enumerate(candid_docs):
123     print(doc_id_to_doc_title.get(doc.doc_id))
124     if i > 5:
125         break
126
127 query
128 'استقلال'
129 for i, doc in enumerate(candid_docs):
130     print(doc_id_to_doc_title.get(doc.doc_id))
131     if i > 5:
132         break
133
134 query
135 'استقلال'
136 for i, doc in enumerate(candid_docs):
137     print(doc_id_to_doc_title.get(doc.doc_id))
138     if i > 5:
139         break
140
141 query
142 'استقلال'
143 for i, doc in enumerate(candid_docs):
144     print(doc_id_to_doc_title.get(doc.doc_id))
145     if i > 5:
146         break
147
148 query
149 'استقلال'
150 for i, doc in enumerate(candid_docs):
151     print(doc_id_to_doc_title.get(doc.doc_id))
152     if i > 5:
153         break
154
155 query
156 'استقلال'
157 for i, doc in enumerate(candid_docs):
158     print(doc_id_to_doc_title.get(doc.doc_id))
159     if i > 5:
160         break
161
162 query
163 'استقلال'
164 for i, doc in enumerate(candid_docs):
165     print(doc_id_to_doc_title.get(doc.doc_id))
166     if i > 5:
167         break
168
169 query
170 'استقلال'
171 for i, doc in enumerate(candid_docs):
172     print(doc_id_to_doc_title.get(doc.doc_id))
173     if i > 5:
174         break
175
176 query
177 'استقلال'
178 for i, doc in enumerate(candid_docs):
179     print(doc_id_to_doc_title.get(doc.doc_id))
180     if i > 5:
181         break
182
183 query
184 'استقلال'
185 for i, doc in enumerate(candid_docs):
186     print(doc_id_to_doc_title.get(doc.doc_id))
187     if i > 5:
188         break
189
190 query
191 'استقلال'
192 for i, doc in enumerate(candid_docs):
193     print(doc_id_to_doc_title.get(doc.doc_id))
194     if i > 5:
195         break
196
197 query
198 'استقلال'
199 for i, doc in enumerate(candid_docs):
200     print(doc_id_to_doc_title.get(doc.doc_id))
201     if i > 5:
202         break
203
204 query
205 'استقلال'
206 for i, doc in enumerate(candid_docs):
207     print(doc_id_to_doc_title.get(doc.doc_id))
208     if i > 5:
209         break
210
211 query
212 'استقلال'
213 for i, doc in enumerate(candid_docs):
214     print(doc_id_to_doc_title.get(doc.doc_id))
215     if i > 5:
216         break
217
218 query
219 'استقلال'
220 for i, doc in enumerate(candid_docs):
221     print(doc_id_to_doc_title.get(doc.doc_id))
222     if i > 5:
223         break
224
225 query
226 'استقلال'
227 for i, doc in enumerate(candid_docs):
228     print(doc_id_to_doc_title.get(doc.doc_id))
229     if i > 5:
230         break
231
232 query
233 'استقلال'
234 for i, doc in enumerate(candid_docs):
235     print(doc_id_to_doc_title.get(doc.doc_id))
236     if i > 5:
237         break
238
239 query
240 'استقلال'
241 for i, doc in enumerate(candid_docs):
242     print(doc_id_to_doc_title.get(doc.doc_id))
243     if i > 5:
244         break
245
246 query
247 'استقلال'
248 for i, doc in enumerate(candid_docs):
249     print(doc_id_to_doc_title.get(doc.doc_id))
250     if i > 5:
251         break
252
253 query
254 'استقلال'
255 for i, doc in enumerate(candid_docs):
256     print(doc_id_to_doc_title.get(doc.doc_id))
257     if i > 5:
258         break
259
260 query
261 'استقلال'
262 for i, doc in enumerate(candid_docs):
263     print(doc_id_to_doc_title.get(doc.doc_id))
264     if i > 5:
265         break
266
267 query
268 'استقلال'
269 for i, doc in enumerate(candid_docs):
270     print(doc_id_to_doc_title.get(doc.doc_id))
271     if i > 5:
272         break
273
274 query
275 'استقلال'
276 for i, doc in enumerate(candid_docs):
277     print(doc_id_to_doc_title.get(doc.doc_id))
278     if i > 5:
279         break
280
281 query
282 'استقلال'
283 for i, doc in enumerate(candid_docs):
284     print(doc_id_to_doc_title.get(doc.doc_id))
285     if i > 5:
286         break
287
288 query
289 'استقلال'
290 for i, doc in enumerate(candid_docs):
291     print(doc_id_to_doc_title.get(doc.doc_id))
292     if i > 5:
293         break
294
295 query
296 'استقلال'
297 for i, doc in enumerate(candid_docs):
298     print(doc_id_to_doc_title.get(doc.doc_id))
299     if i > 5:
300         break
301
302 query
303 'استقلال'
304 for i, doc in enumerate(candid_docs):
305     print(doc_id_to_doc_title.get(doc.doc_id))
306     if i > 5:
307         break
308
309 query
310 'استقلال'
311 for i, doc in enumerate(candid_docs):
312     print(doc_id_to_doc_title.get(doc.doc_id))
313     if i > 5:
314         break
315
316 query
317 'استقلال'
318 for i, doc in enumerate(candid_docs):
319     print(doc_id_to_doc_title.get(doc.doc_id))
320     if i > 5:
321         break
322
323 query
324 'استقلال'
325 for i, doc in enumerate(candid_docs):
326     print(doc_id_to_doc_title.get(doc.doc_id))
327     if i > 5:
328         break
329
330 query
331 'استقلال'
332 for i, doc in enumerate(candid_docs):
333     print(doc_id_to_doc_title.get(doc.doc_id))
334     if i > 5:
335         break
336
337 query
338 'استقلال'
339 for i, doc in enumerate(candid_docs):
340     print(doc_id_to_doc_title.get(doc.doc_id))
341     if i > 5:
342         break
343
344 query
345 'استقلال'
346 for i, doc in enumerate(candid_docs):
347     print(doc_id_to_doc_title.get(doc.doc_id))
348     if i > 5:
349         break
350
351 query
352 'استقلال'
353 for i, doc in enumerate(candid_docs):
354     print(doc_id_to_doc_title.get(doc.doc_id))
355     if i > 5:
356         break
357
358 query
359 'استقلال'
360 for i, doc in enumerate(candid_docs):
361     print(doc_id_to_doc_title.get(doc.doc_id))
362     if i > 5:
363         break
364
365 query
366 'استقلال'
367 for i, doc in enumerate(candid_docs):
368     print(doc_id_to_doc_title.get(doc.doc_id))
369     if i > 5:
370         break
371
372 query
373 'استقلال'
374 for i, doc in enumerate(candid_docs):
375     print(doc_id_to_doc_title.get(doc.doc_id))
376     if i > 5:
377         break
378
379 query
380 'استقلال'
381 for i, doc in enumerate(candid_docs):
382     print(doc_id_to_doc_title.get(doc.doc_id))
383     if i > 5:
384         break
385
386 query
387 'استقلال'
388 for i, doc in enumerate(candid_docs):
389     print(doc_id_to_doc_title.get(doc.doc_id))
390     if i > 5:
391         break
392
393 query
394 'استقلال'
395 for i, doc in enumerate(candid_docs):
396     print(doc_id_to_doc_title.get(doc.doc_id))
397     if i > 5:
398         break
399
400 query
401 'استقلال'
402 for i, doc in enumerate(candid_docs):
403     print(doc_id_to_doc_title.get(doc.doc_id))
404     if i > 5:
405         break
406
407 query
408 'استقلال'
409 for i, doc in enumerate(candid_docs):
410     print(doc_id_to_doc_title.get(doc.doc_id))
411     if i > 5:
412         break
413
414 query
415 'استقلال'
416 for i, doc in enumerate(candid_docs):
417     print(doc_id_to_doc_title.get(doc.doc_id))
418     if i > 5:
419         break
420
421 query
422 'استقلال'
423 for i, doc in enumerate(candid_docs):
424    
```



Category = health

Query = کرونا



```
24 candid_docs.sort(key=lambda x: x.tern_weight_dict.get(tern_id))
25
26 print('select from posting:')
27 candid_docs = []
28 for doc_id, posting in query_posting_list.posting_list.items():
29     doc_vertx = obj.doc_vertex.get(doc_id)
```

```
cat
'health'
for i, doc in enumerate(candid_docs):
    print(doc_id_to_doc_title.get(doc.doc_id))
    if i > 5:
        break
```

<https://www.iana.ir/news/99010200789/> ثبت یک روز بدون کرونا بی جدید در استان سمنان

<https://www.iana.ir/news/99010200594/> ثبت اطلاعات خود به سامانه خود مراقبتی سلامت مراجعه

<https://www.iana.ir/news/99010401017/> برای تطخیم راست از دروغ راه نیست به جز با فن راسگوبان درمان

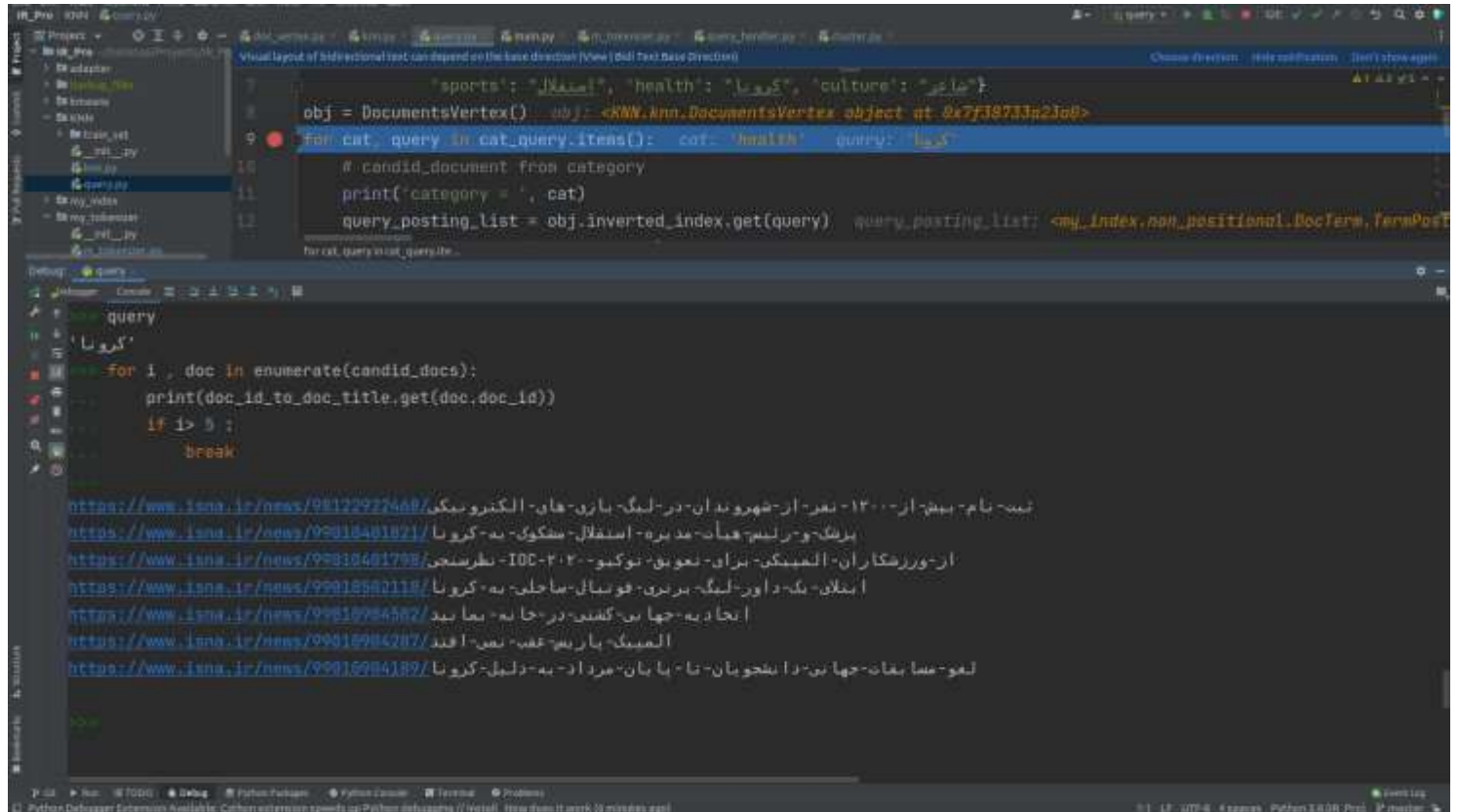
<https://www.iana.ir/news/99010301191/> چگونه بر استریم ناشی از همه گیری و بروز کرونا غلبه کنیم

<https://www.iana.ir/news/99010703339/> نزاع بین ۲۸۳۳۰ میلیارد تومان دیگر توسط دولت به نظام سلامت

<https://www.iana.ir/news/99010705106/> بهمدی حال عمومی رئیس سازمان اورژانس کشور پس از ابتلا به کرونا

<https://www.iana.ir/news/99010703276/> با همه بین از ۴۰۰ سازمان مردم نهاد به سازمان ملل برای لغو تحریم های

Query = کرونا

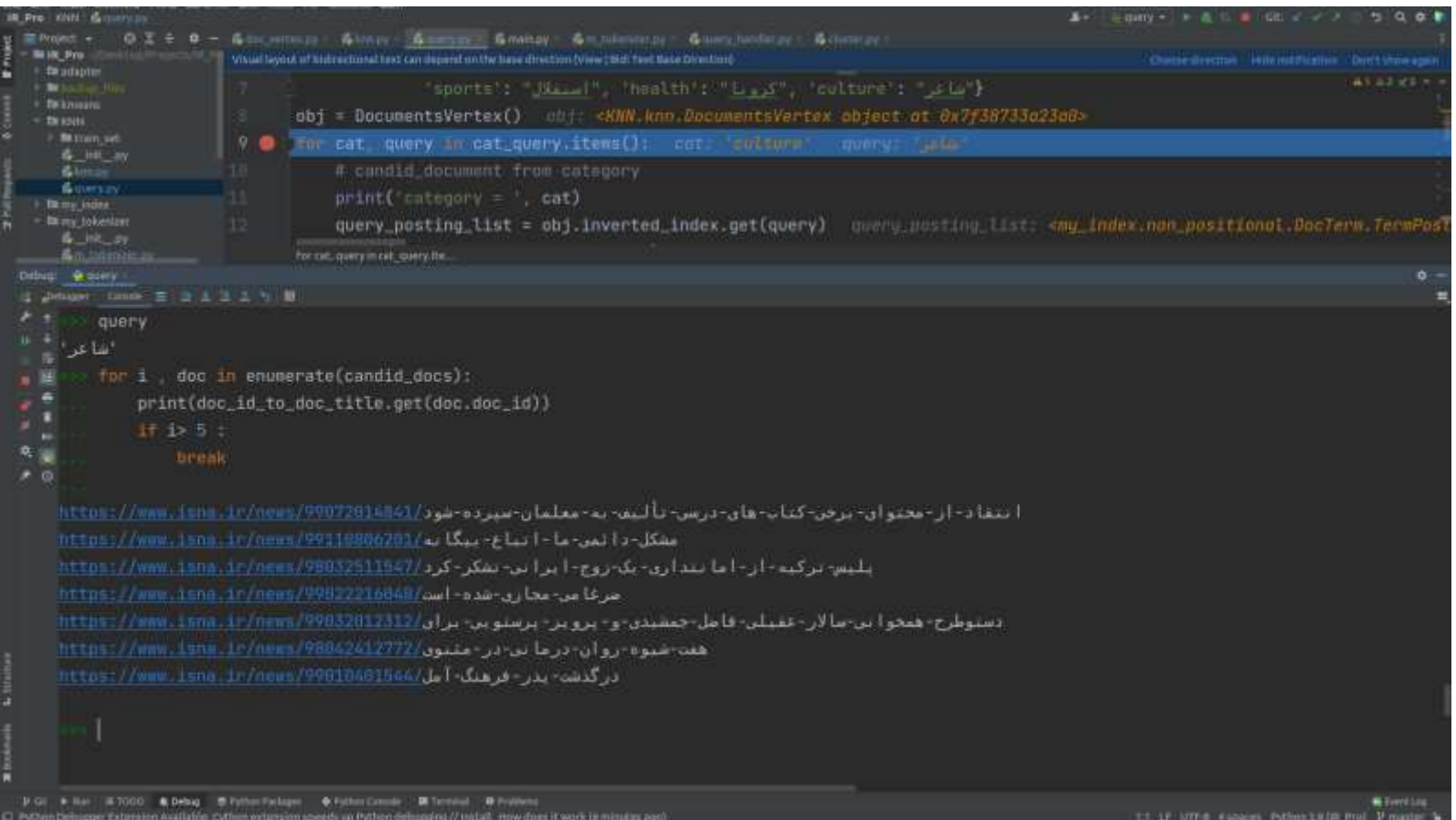


Query = شاعر

[illegible]

Without category

Query = شاعر



```
7     "sports": "اسفند", "health": "کرونا", "culture": "شاعر"}
8     obj = DocumentsVertex()  obj: <KNN.knn.DocumentsVertex object at 0x7f38733a23a8>
9     for cat, query in cat_query.items():  cat: 'culture'  query: 'شاعر'
10    # candid document from category
11    print('category = ', cat)
12    query_posting_list = obj.inverted_index.get(query)  query_posting_list: <my_index.non_positional.DocTerm.TermPost
for cat, query in cat_query.items():
```

```
>>> query
'شاعر'
>>> for i, doc in enumerate(candid_docs):
...     print(doc_id_to_doc_title.get(doc.doc_id))
...     if i > 5:
...         break
...
https://www.isna.ir/news/99072816841/انتقاد-از-محتوای-برخی-کتاب-های-درس-تألیف-به-معلمان-سپرده-شود
https://www.isna.ir/news/99110806203/مشکل-دا-نمن-ما-انواع-بیگانه
https://www.isna.ir/news/98032911947/پلیس-ترکیه-از-اما-نداری-یک-روح-ا-برا-نی-بشکر-کرد
https://www.isna.ir/news/99822216848/جرعا-می-مجا-ی-شده-است
https://www.isna.ir/news/99032017312/دستورطرح-فمخوا-بی-سالار-تغییر-فاحل-جمشیدی-و-برو-بر-پرسوبی-برای
https://www.isna.ir/news/98042417772/هفت-شبهه-روان-درما-بی-در-مثنوی
https://www.isna.ir/news/99010401544/درگذشت-بدر-فرهنگ-آمل
```

در نتیجه اگر کویری در کلاس مشخص خودش بررسی شود داکيومنت های مرتبط تری بازگردانده میشوند و این هم چنین سرعت بیشتری در بازیابی هم دارد