

به نام خدا

گزارش فاز دوم پروژه بازیابی اطلاعات"

محمدرضا قادری ۹۶۲۷۰۵۷

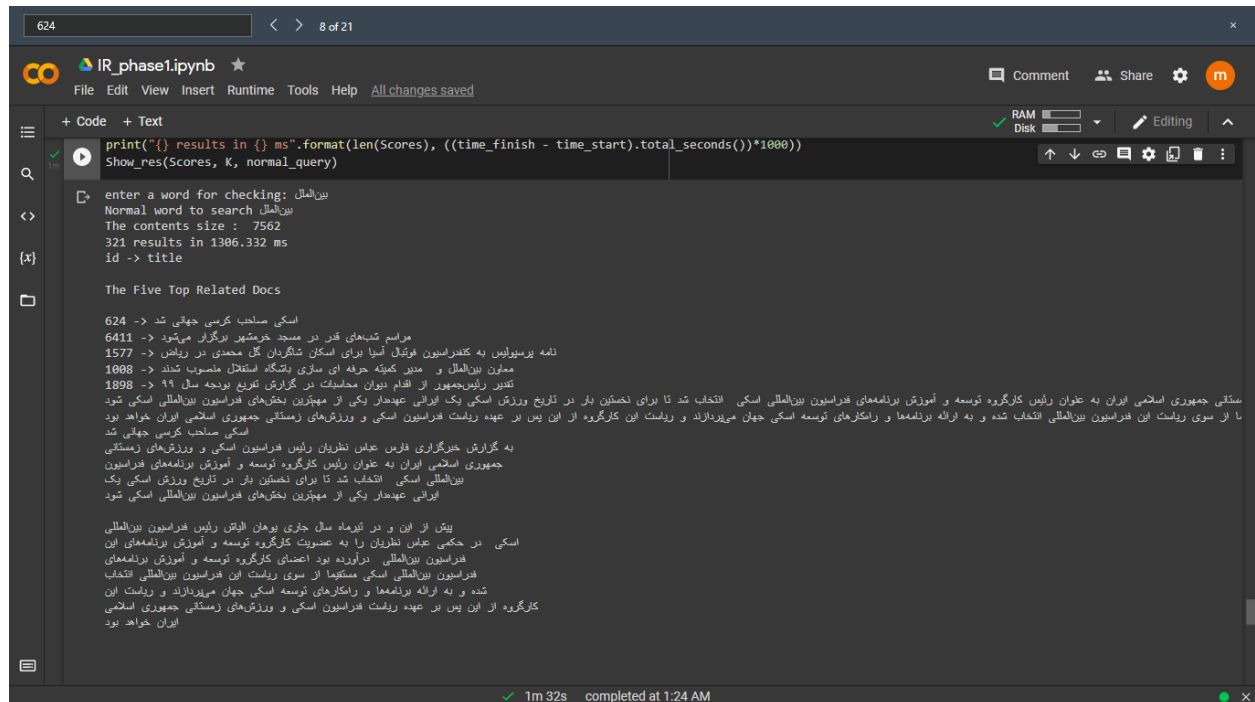
در این فاز ما به جای استفاده از Posting list مربوط به هر کلمه برای ارتباط درخواست‌ها استفاده می‌کردیم ولی در اینجا می‌خواهیم با استفاده از tfidf برای شاخص استفاده کنیم.

ما برای هر کلمه برای هر سند tf , idf , $tfidf$ را محاسبه می‌کنیم و برای محاسبه شباهت نیز از روش کسینوسی استفاده خواهیم کرد پس همین جا که $tfidf$ حساب می‌کنیم طولش را نیز برای هر سند استفاده می‌کنیم. حالا ما میتوانیم با استفاده از همین $tfidf$ ها میتوانیم Champion List تشکیل دهیم، در اینجا ما Index Elimination لازم نداریم چرا که از روش TaaT استفاده می‌کنیم خودش این قضیه رو برای ما حل می‌کند. برای Champion list نیز هم چندین روش وجود دارد یکی از روش‌ها استفاده از یک ترشهولد برای اینکه در این لیست از اون کلمات استفاده کنیم و روش دیگه نیز می‌توانیم با سورت کردن همه لیست‌ها بر حسب $tfidf$ در ابتدا و سپس بر حسب شماره داک می‌تونیم میتواند برای آن استفاده کنیم ازش برای همین که حساب کردن ترشهولد نیازمند این است که کلیه مقادیر رو بشناسیم اعم از میانگین و ماکسیمم و شاید هم مینیمم که بتوانیم یک محدوده‌ای برای اینکه $tfidf$ مشخص کنیم پس ما از روش اول استفاده می‌کنیم. گزارش از ما خواسته که با این کار ابتدا Query خواسته شده را با این روش با doc ها مقایسه کنیم و شباهت رو بدست بیاریم و سپس با روش اول که در فاز اول داشتیم مقایسه انجام دهیم. از همان کلمات که در فاز اول استفاده شده بود برای مقایسه استفاده شد.

۱) پرسمان‌های برای tfidf

الف) یک پرسمان از کلمات ساده و متداول تک کلمه‌ای

بین الملل



The screenshot shows a Jupyter Notebook titled "IR_phase1.ipynb". The code cell contains a Python script that prints the results of a TF-IDF search for the word "بین‌الملل". The output shows the word size, the number of results (321), and the top five related documents. The documents are listed with their IDs and titles.

```
print("{} results in {} ms".format(len(Scores), ((time_finish - time_start).total_seconds()*1000))
Show_res(Scores, K, normal_query)
```

بین‌الملل
Normal word to search
The contents size : 7562
321 results in 1306.332 ms
id -> title

The Five Top Related Docs

624 -> اسکی صاحب کرسی جهانی شد ->
6411 -> مراسم شب‌های قدر در مسجد خرمشهر برگزار می‌شود ->
1577 -> تلمه پرسپولیس به کنفدراسیون فوتبال آسیا برای اسکان شگردان گل محمدی در ریاض ->
1008 -> معاون بین‌الملل و مدیر کمیته حرفه ای سوئی دانشگاه استقلال منصوب شدند ->
1898 -> تقدیر رئیس‌جمهور از قدام دیوان محاسبات در گزارش تفریح بونجه سال ۹۹ ->
سنتی جمهوری اسلامی ایران به عنوان رئیس کارگروه توسعه و آموزش برنامه‌های فراسیون بین‌المللی اسکی انتخاب شد تا برای نخستین بار در تاریخ ورزش اسکی جهان می‌پردازد و ریاست این کارگروه از این پس بر عهده ریاست فراسیون اسکی و ورزش‌های زمستانی جمهوری اسلامی ایران خواهد بود
اسکی صاحب کرسی جهانی شد
به گزارش خبرنگاری فارس عباس نظریان رئیس فراسیون اسکی و ورزش‌های زمستانی جمهوری اسلامی ایران به عنوان رئیس کارگروه توسعه و آموزش برنامه‌های فراسیون بین‌المللی اسکی انتخاب شد تا برای نخستین بار در تاریخ ورزش اسکی یک ایرانی عهده‌دار یکی از مهمترین بخش‌های فراسیون بین‌المللی اسکی شود

بیان از این و در تیرماه سال جاری یوهان الوتش رئیس فراسیون بین‌المللی اسکی در حکمی عباس نظریان را به عضویت کارگروه توسعه و آموزش برنامه‌های این فراسیون بین‌المللی درآورده بود اعضای کارگروه توسعه و آموزش برنامه‌های فراسیون بین‌المللی اسکی مستکماً از سوی ریاست این فراسیون بین‌المللی انتخاب شده و به ارائه برنامه‌ها و رانک‌های توسعه اسکی جهان می‌پردازد و ریاست این کارگروه از این پس بر عهده ریاست فراسیون اسکی و ورزش‌های زمستانی جمهوری اسلامی ایران خواهد بود

با توجه به اینکه برای ما موثر هست اندازه متن می‌بینیم در بین متن‌های بدست آمده تاثیر خودش را گذاشته هست و بله این سند مرتبط هست و تعداد کلمات منحصر به فرد زیادی به کار رفته در کوئری ما هست به نسبت اندازه کل سند.

ب)یک پرسمان از عبارات ساده و متداول چندکلمه‌ای

دانشگاه امیرکبیر

CO

IR phase1.ipynb

★

File Edit View Insert Runtime Tools Help All changes saved

Comment Share Settings m

+ Code + Text

RAM 100% Disk 100%

Editing

enter a word for checking: دانشگاه امیرکبیر

Normal word to search دانشگاه امیرکبیر

756 results in 88.7610000000001 ms

id -> title

The Five Top Related Docs

2793

یاد برای ثبت نش دانشگاهیان در دوران دفاع مقدس کار تحصیلاتی صورت گیرد ->

2792

امروز محیط دانشگاههای ما عرصه دفاع مقدس است ->

7520

اعضای گرامی درجه یک هنری به سپید بویانگی ->

7236

برگزاری یادبودی برای محمدرسول رجایی ->

5021

دقرچه راهمادی آزمون استعدادی دانشگاهها برای بار چهارم اصلاح شد/نمید مجد مهلت ثبتنام ->

سوع خیلی مهم هست که بدایم دانشگاهیان دانشگاهیان در تاریخ انقلاب و دوران دفاع مقدس چه نقشی داشتند و جدای از آن وظیفه اصلی که از آنها انتظار میرفت و تحصیل در دانشگاه است چه اقدامی برای جامعه و انقلاب داشتند است و ما همواره بر این مسئله تاکید داریم ولی در کنار این موضوع دانشجویی که فارغ التحصیل میشود هر چقدر با مسائل اجتماعی سیاسی فرهنگی جامعه اشتغال باشد میتواند نقش مؤثرتری را در کنار کارهای علمی در جامعه ایفا کند که با حکومت طاغوت مبارزه کردند از همین دانشگاهها بودند دانشجویان و جوانان در دوران انقلاب نقش ویژه ای داشتند بسیاری از قلمها و مستندات را که نگاه میکنیم میبینیم اکثر جوانان نقش داشتند حفظ انقلاب هم بر دوش جوانان است بر ادامه با اشاره به نقش دانشجویان و جوانان در تشکیل نیروهای بسیج بسیج و سیاه گت وقتی خاطرات فراموشان بزرگ را مشاهده میکنیم میبینیم که آنها دانشجویانی بودند که به جبهه های دفاع مقدس رفتند و در مقابل دشمن ایستادند محمندی افزود اگر بخواهیم در مورد نقش دانشجوی و دانشگاه در جبهه های جنگ و در دوران دفاع مقدس صحبت کنیم نواز به یک کار تحصیلاتی است که توسیه میکنم به این کار پرداخته شود وی افزود بیش از شهید دوران دفاع مقدس از دانشگاه امیرکبیر بودند یعنی در زمانی که دانشگاه حدود دانشجو داشت دانشگاهیان این دانشگاه این گونه در خدمت جبهه و جنگ بودند در دوران دفاع مقدس جهت انتقال به نسل جدید بسیار اهمیت دارد البته در بخش های مختلف این اتفاق افتاده و اگر یک تیم محقق بپشتند و حمایت دانشگاهیان در زمان جنگ تحصیلی را ثبت کنند کار بسیار خوبی خوبی انجام خواهد شد یاد برای ثبت نش دانشگاهیان در دوران دفاع مقدس کار تحصیلاتی صورت گیرد به گزارش خبرنگار شکل های دانشگاهی برگزاری فارس سوادمحند محمندی رئیس دانشگاه سمنانی امیرکبیر دانشگاه سمنانی امیرکبیر

تشنه ای که به سمت نهاد نمایندگی مقام معظم رهبری در دانشگاهها به مناسبت فرارسیدن هفته دفاع مقدس برگزار شد دربرهه نش آموزش عالی در حیطه و انگیاریی خاطرات دوران دفاع مقدس و یاد و خاطره شهدای جنگ تحصیلی گت این موضوع خیلی مهم هست که بدایم دانشگاهیان دانشگاهیان در تاریخ انقلاب و

دوران دفاع مقدس چه نقشی داشتند و جدای از آن وظیفه اصلی که از آنها انتظار میرفت و تحصیل در دانشگاه است چه اقدامی برای جامعه و انقلاب داشتند

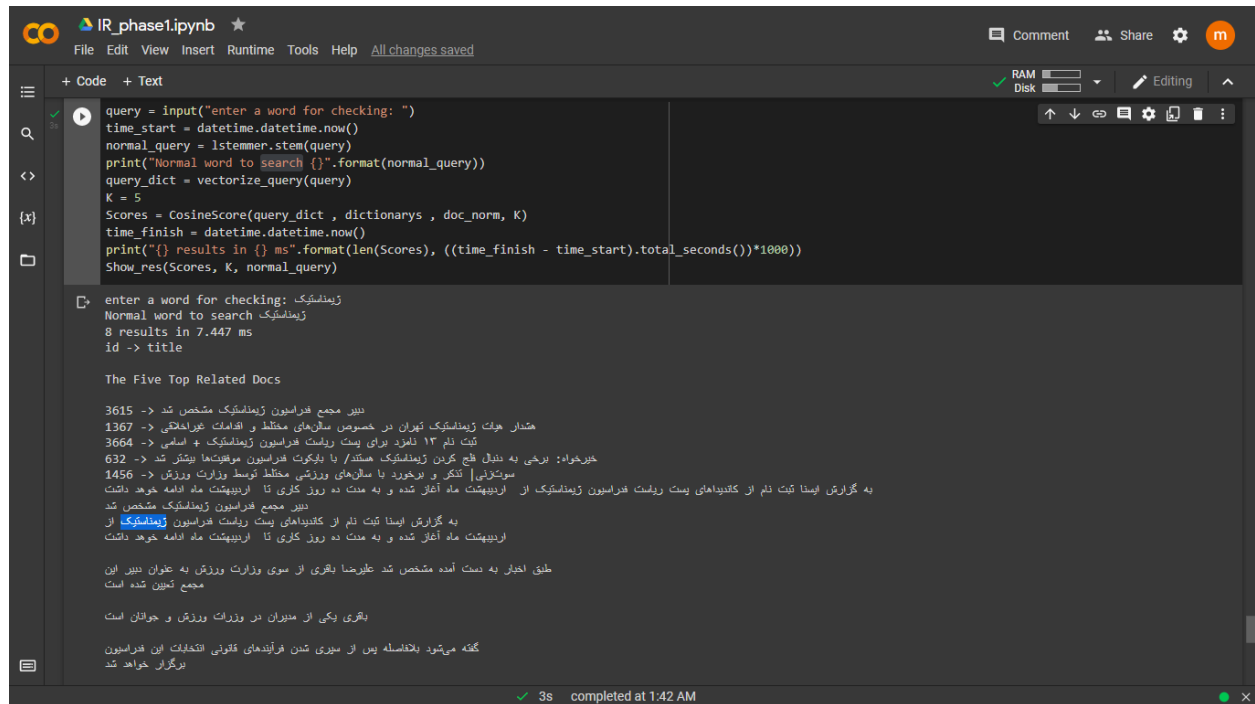
مقدمی ترویج کرد برخی فکر میکنند وقتی وارد دانشگاه میشوند فقط باید درس بخوانند که البته این کار اصلی دانشجو است و ما همواره بر این مسئله تاکید داریم ولی در کنار این موضوع دانشجویی که فارغ التحصیل میشود هر چقدر با مسائل اجتماعی سیاسی فرهنگی جامعه اشتغال باشد میتواند نقش مؤثرتری را در کنار کارهای علمی در جامعه ایفا کند

وی افزود وقتی دانشجو وارد دانشگاه میشود باید بداند دانشگاه و دانشگاهی که در

8s completed at 1:34 AM

بله این نیز مرتبط هست خبر حاوی تعداد خوبی کلمه دانشگاه امیرکبیر هست و سند نیز مربوط به سخنان رییس دانشگاه می باشد.

پ) یک پرسمان دشوار و کم تکرار تک‌کلمه‌ای
ژیمناستیک



```
query = input("enter a word for checking: ")
time_start = datetime.datetime.now()
normal_query = lstemmer.stem(query)
print("Normal word to search {}".format(normal_query))
query_dict = vectorize_query(query)
K = 5
Scores = CosineScore(query_dict, dictionarys, doc_norm, K)
time_finish = datetime.datetime.now()
print("{} results in {} ms".format(len(Scores), ((time_finish - time_start).total_seconds()*1000))
Show_res(Scores, K, normal_query)
```

enter a word for checking: ژیمناستیک
ژیمناستیک
Normal word to search ژیمناستیک
8 results in 7.447 ms
id -> title

The Five Top Related Docs

دبیر مجمع فدراسیون ژیمناستیک مشخص شد -> 3615
مدار هات ژیمناستیک تهران در خصوص سال‌های مختلف و اقدامات غرضی -> 1367
ثبت نام ۱۳ نامزد برای پست ریاست فدراسیون ژیمناستیک + اسامی -> 3664
خبرخواه: برخی به دنبال فتح کردن ژیمناستیک هستند/ با بایکوت فدراسیون موافقت‌ها بیشتر شد -> 632
سوت‌زنی | تذکر و برخورد با سال‌های ورزشی مختلف توسط وزارت ورزش -> 1456
به گزارش ایسنا ثبت نام از کفپادهای پست ریاست فدراسیون ژیمناستیک از اردیبهشت ماه آغاز شده و به مدت ده روز کاری تا اردیبهشت ماه ادامه خواهد داشت
دبیر مجمع فدراسیون ژیمناستیک مشخص شد
به گزارش ایسنا ثبت نام از کفپادهای پست ریاست فدراسیون ژیمناستیک از اردیبهشت ماه آغاز شده و به مدت ده روز کاری تا اردیبهشت ماه ادامه خواهد داشت
طبق اخبار به دست آمده مشخص شد علیرضا باقری از سوی وزارت ورزش به عنوان دبیر این مجمع تعیین شده است
باقری یکی از منیران در وزارت ورزش و جوانان است
گفته می‌شود بلافاصله پس از سبزی شدن فرآیندهای قانونی انتخابات این فدراسیون برگزار خواهد شد

3s completed at 1:42 AM

باز هم مرتبط است خبر منتشر شده درباره فدراسیون ژیمناستیک که در باره تغییر دبیر فدراسیون هست
در اینجا باز هم تاثیر *idf* رو مشاهده می‌کنیم که به خاطر کم بودن متن و تکرار یک باره این متن
مرتبط شده است.

ت) یک پرسمان دشوار و کم تکرار چندکلمه‌ای

واکسن آسترزنکا

IR phase1.ipynb

File Edit View Insert Runtime Tools Help All changes saved

+ Code + Text

enter a word for checking: واکسن آسترزنکا
واکسن آسترزنکا
Normal word to search
594 results in 50.772 ms
id -> title

The Five Top Related Docs

آمار تکلیفی واکسیناسیون کرونا در ایران تا ۲۶ مردادماه -> 5889
آمار تکلیفی واکسیناسیون کرونا در ایران تا ۲۶ مردادماه -> 5896
اعلام آمار تکلیفی واکسیناسیون کرونا تا صبح امروز -> 5936
آمار تکلیفی واکسیناسیون کرونا در کشور تا ۲۱ مرداد -> 5864
آمار تکلیفی واکسیناسیون کرونا در کشور تا ۲۱ مرداد -> 5853
به گزارش ایسنا دکتر کیانوش جهانپور رئیس مرکز روابط عمومی و اطلاع رسانی وزارت بهداشت آمار تکلیفی واکسن مسرفی کرونا در کشور از ابتدای برنامه ملی واکسیناسیون علوه گوید تا مرداد را اعلام کرد
سینوفارم نوبت اول
سینوفارم نوبت دوم
آسترزنکا نوبت اول
آسترزنکا نوبت دوم
برکت نوبت اول
برکت نوبت دوم
اسپوتنیک نوبت اول
اسپوتنیک نوبت دوم
بهارات نوبت اول
بهارات نوبت دوم
آمار تکلیفی واکسیناسیون کرونا در ایران تا ۲۶ مردادماه
به گزارش ایسنا دکتر کیانوش جهانپور رئیس مرکز روابط عمومی و اطلاع رسانی وزارت بهداشت آمار تکلیفی واکسن مسرفی کرونا در کشور از ابتدای برنامه ملی واکسیناسیون علوه گوید تا مرداد را اعلام کرد
سینوفارم نوبت اول
سینوفارم نوبت دوم
آسترزنکا نوبت اول
آسترزنکا نوبت دوم

3s completed at 1:47 AM

بله این نیز مرتبط هست که به نوبت واکسن آسترزنکا مربوط می‌باشد.

برای مقایسه فاز ۱ و ۲

حالت ب)

فاز ۲

```
IR_phase1.ipynb
File Edit View Insert Runtime Tools Help All changes saved

+ Code + Text

enter a word for checking: دانشگاه امیرکبیر
Normal word to search دانشگاه امیرکبیر
756 results in 88.76100000000001 ms
id -> title

The Five Top Related Docs

2793 -> باید برای ثبت نش در دوران دفاع مقدس کار تحقیقاتی صورت گیرد ->
2792 -> امروز محیط دانشگاههای ما عرصه دفاع مقدس است ->
7520 -> اعطای گرامی درجه یک هنری به سید بیژانکی ->
7236 -> برگزاری یلپودی برای محسنرور رجایی ->
5021 -> دفرجه راهمائی آزمون استعدای دانشگاهها برای بار چهارم اصلاح شد/تمدید مجدد مهلت استخدام ->
منوع خیلی مهم هست که بدایم دانشگاهیان دانشگاهیان در تاریخ انقلاب و دوران دفاع مقدس چه نقشی داشتند و جدای از آن وظیفه اصلی که از آنها انتظار میرفت و تحصیل در دانشگاه است چه اقداماتی برای جامعه و انقلاب داشتند
است و ما همواره بر این مسئله تاکید داریم ولی در کنار این موضوع دانشجوی که فارغ التحصیل میشود هر چند با مسائل اجتماعی سیاسی فرهنگی جامعه اشتغال باشد میتواند نقش مؤثرتری را در کنار کارهای علمی در جامعه ایفا کند
که با حکومت طاغوت مبارزه کردند از همین دانشگاهها بودند دانشجویان و جوانان در دوران انقلاب نقش ویژه داشتند بسیاری از قلمها و مستندات را که نگاه میکنیم میبینیم اکثر آنها دانشجویانی بودند که به جبهه های دفاع مقدس رفتند و در مقابل دشمن ایستادند
در ادامه با اشاره به نقش دانشجویان و جوانان در تشکیل نیروهای بسیج بسیج و سیاه گت وقتی خطرات فرماندهان بزرگ را مشاهده میکنیم میبینیم که آنها دانشجویانی بودند که به جبهه های دفاع مقدس رفتند و در مقابل دشمن ایستادند
معمدی افزود اگر بخواهیم در مورد نقش دانشجو و دانشگاه در جبهه های جنگ و در دوران دفاع مقدس صحبت کنیم نیاز به یک کار تحقیقاتی است که توسیه میکنم به این کار پرداخته شود
وی افزود بیش از شپید دوران دفاع مقدس از دانشگاه امیرکبیر بودند پسلی در زمانی که دانشگاه حدود دانشجو داشت دانشگاهیان این دانشگاه این گروه در خدمت جبهه و جنگ بودند
در دوران دفاع مقدس جهت انتقال به نسل جدید بسوی اهمیت دارد البته در بخشهای مختلف این اتفاق افتاده و اگر یک نیم محقق بشنایند و خدمات دانشگاهیان در زمان جنگ تحصیلی را ثبت کنند کار بسوی خوبی خوبی انجام خواهد شد
باید برای ثبت نش دانشگاهیان در دوران دفاع مقدس کار تحقیقاتی صورت گیرد
به گزارش خبرنگار تشکلهای دانشگاهی خبرگزاری فارس سیداحمد معمدی رئیس دانشگاه
صلحی امیرکبیر دانشگاه صنعتی امیرکبیر در
تشستی که به همت نهاد نمایندگی مقام معظم رهبری در دانشگاهها به مناسبت فرارسیدن
هفته دفاع مقدس برگزار شد درباره نقش آموزش عالی در حفظ و نگهداری خاطرات دوران
دفاع مقدس و یاد و خاطره شهدای جنگ تحصیلی گفت این موضوع خیلی مهم هست که بدایم
دانشگاهیان دانشگاهیان در تاریخ انقلاب و
دوران دفاع مقدس چه نقشی داشتند و جدای از آن وظیفه اصلی که از آنها انتظار
میرفت و تحصیل در دانشگاه است چه اقداماتی برای جامعه و انقلاب داشتند
معمدی تلمریج کرد برخی فکر میکنند وقتی وارد دانشگاه میشوند فقط باید درس
بخوانند که البته این کار اصلی دانشجو است و ما همواره بر این مسئله تاکید داریم
ولی در کنار این موضوع دانشجوی که فارغ التحصیل میشود هر چند با مسائل اجتماعی
سیاسی فرهنگی جامعه اشتغال باشد میتواند نقش مؤثرتری را در کنار کارهای علمی در
جامعه ایفا کند
وی افزود وقتی دانشجو وارد دانشگاه میشود باید بداند دانشگاه و دانشگاهی که در
8s completed at 1:34 AM
```

فاز ۱

```
IR_phase1.ipynb
File Edit View Insert Runtime Tools Help

+ Code + Text

بهرات لوبت دوم [221]

[214] query_one_word()

query_multi()

enter words for checking: دانشگاه امیرکبیر
دانشگاه
2621
[[362, 1, [195]], [390, 10, [18, 21, 76, 206, 209, 238, 263, 342, 362, 374]], [393, 1, [18]], [552, 2, [108, 167]], [560, 1, [80]], [724, 3, [89, 133, 136]], [775, 1, [1753, 1, [4038]], [1959, 1, [574]], [2131, 1, [511]], [2709, 1, [436]], [2792, 4, [18, 182, 223, 302]], [2793, 4, [12, 15, 231, 316]], [4580, 1, [95]], [5021, 1, [2621, [[362, 1, [195]], [390, 10, [18, 21, 76, 206, 209, 238, 263, 342, 362, 374]], [393, 1, [18]], [552, 2, [108, 167]], [560, 1, [80]], [724, 3, [89, 133, 136]], [775, 1, [1753, 1, [4038]], [1959, 1, [574]], [2131, 1, [511]], [2709, 1, [436]], [2792, 4, [18, 182, 223, 302]], [2793, 4, [12, 15, 231, 316]], [4580, 1, [95]], [5021, 1, [2621, [[362, 1, [195]], [390, 10, [18, 21, 76, 206, 209, 238, 263, 342, 362, 374]], [393, 1, [18]], [552, 2, [108, 167]], [560, 1, [80]], [724, 3, [89, 133, 136]], [775, 1, [1753, 1, [4038]], [1959, 1, [574]], [2131, 1, [511]], [2709, 1, [436]], [2792, 4, [18, 182, 223, 302]], [2793, 4, [12, 15, 231, 316]], [4580, 1, [95]], [5021, 1, [2621, [[362, 1, [195]], [390, 10, [18, 21, 76, 206, 209, 238, 263, 342, 362, 374]], [393, 1, [18]], [552, 2, [108, 167]], [560, 1, [80]], [724, 3, [89, 133, 136]], [775, 1, [1753, 1, [4038]], [1959, 1, [574]], [2131, 1, [511]], [2709, 1, [436]], [2792, 4, [18, 182, 223, 302]], [2793, 4, [12, 15, 231, 316]], [4580, 1, [95]], [5021, 1, [2621, [[362, 1, [195]], [390, 10, [18, 21, 76, 206, 209, 238, 263, 342, 362, 374]], [393, 1, [18]], [552, 2, [108, 167]], [560, 1, [80]], [724, 3, [89, 133, 136]], [775, 1, [1753, 1, [4038]], [1959, 1, [574]], [2131, 1, [511]], [2709, 1, [436]], [2792, 4, [18, 182, 223, 302]], [2793, 4, [12, 15, 231, 316]], [4580, 1, [95]], [5021, 1, [2621, [[362, 1, [195]], [390, 10, [18, 21, 76, 206, 209, 238, 263, 342, 362, 374]], [393, 1, [18]], [552, 2, [108, 167]], [560, 1, [80]], [724, 3, [89, 133, 136]], [775, 1, [1753, 1, [4038]], [1959, 1, [574]], [2131, 1, [511]], [2709, 1, [436]], [2792, 4, [18, 182, 223, 302]], [2793, 4, [12, 15, 231, 316]], [4580, 1, [95]], [5021, 1, [2621, [[362, 1, [195]], [390, 10, [18, 21, 76, 206, 209, 238, 263, 342, 362, 374]], [393, 1, [18]], [552, 2, [108, 167]], [560, 1, [80]], [724, 3, [89, 133, 136]], [775, 1, [1753, 1, [4038]], [1959, 1, [574]], [2131, 1, [511]], [2709, 1, [436]], [2792, 4, [18, 182, 223, 302]], [2793, 4, [12, 15, 231, 316]], [4580, 1, [95]], [5021, 1, [2621, [[362, 1, [195]], [390, 10, [18, 21, 76, 206, 209, 238, 263, 342, 362, 374]], [393, 1, [18]], [552, 2, [108, 167]], [560, 1, [80]], [724, 3, [89, 133, 136]], [775, 1, [1753, 1, [4038]], [1959, 1, [574]], [2131, 1, [511]], [2709, 1, [436]], [2792, 4, [18, 182, 223, 302]], [2793, 4, [12, 15, 231, 316]], [4580, 1, [95]], [5021, 1, [2621, [[362, 1, [195]], [390, 10, [18, 21, 76, 206, 209, 238, 263, 342, 362, 374]], [393, 1, [18]], [552, 2, [108, 167]], [560, 1, [80]], [724, 3, [89, 133, 136]], [775, 1, [1753, 1, [4038]], [1959, 1, [574]], [2131, 1, [511]], [2709, 1, [436]], [2792, 4, [18, 182, 223, 302]], [2793, 4, [12, 15, 231, 316]], [4580, 1, [95]], [5021, 1, [2621, [[362, 1, [195]], [390, 10, [18, 21, 76, 206, 209, 238, 263, 342, 362, 374]], [393, 1, [18]], [552, 2, [108, 167]], [560, 1, [80]], [724, 3, [89, 133, 136]], [775, 1, [1753, 1, [4038]], [1959, 1, [574]], [2131, 1, [511]], [2709, 1, [436]], [2792, 4, [18, 182, 223, 302]], [2793, 4, [12, 15, 231, 316]], [4580, 1, [95]], [5021, 1, [2621, [[362, 1, [195]], [390, 10, [18, 21, 76, 206, 209, 238, 263, 342, 362, 374]], [393, 1, [18]], [552, 2, [108, 167]], [560, 1, [80]], [724, 3, [89, 133, 136]], [775, 1, [1753, 1, [4038]], [1959, 1, [574]], [2131, 1, [511]], [2709, 1, [436]], [2792, 4, [18, 182, 223, 302]], [2793, 4, [12, 15, 231, 316]], [4580, 1, [95]], [5021, 1, [2621, [[362, 1, [195]], [390, 10, [18, 21, 76, 206, 209, 238, 263, 342, 362, 374]], [393, 1, [18]], [552, 2, [108, 167]], [560, 1, [80]], [724, 3, [89, 133, 136]], [775, 1, [1753, 1, [4038]], [1959, 1, [574]], [2131, 1, [511]], [2709, 1, [436]], [2792, 4, [18, 182, 223, 302]], [2793, 4, [12, 15, 231, 316]], [4580, 1, [95]], [5021, 1, [2621, [[362, 1, [195]], [390, 10, [18, 21, 76, 206, 209, 238, 263, 342, 362, 374]], [393, 1, [18]], [552, 2, [108, 167]], [560, 1, [80]], [724, 3, [89, 133, 136]], [775, 1, [1753, 1, [4038]], [1959, 1, [574]], [2131, 1, [511]], [2709, 1, [436]], [2792, 4, [18, 182, 223, 302]], [2793, 4, [12, 15, 231, 316]], [4580, 1, [95]], [5021, 1, [2621, [[362, 1, [195]], [390, 10, [18, 21, 76, 206, 209, 238, 263, 342, 362, 374]], [393, 1, [18]], [552, 2, [108, 167]], [560, 1, [80]], [724, 3, [89, 133, 136]], [775, 1, [1753, 1, [4038]], [1959, 1, [574]], [2131, 1, [511]], [2709, 1, [436]], [2792, 4, [18, 182, 223, 302]], [2793, 4, [12, 15, 231, 316]], [4580, 1, [95]], [5021, 1, [2621, [[362, 1, [195]], [390, 10, [18, 21, 76, 206, 209, 238, 263, 342, 362, 374]], [393, 1, [18]], [552, 2, [108, 167]], [560, 1, [80]], [724, 3, [89, 133, 136]], [775, 1, [1753, 1, [4038]], [1959, 1, [574]], [2131, 1, [511]], [2709, 1, [436]], [2792, 4, [18, 182, 223, 302]], [2793, 4, [12, 15, 231, 316]], [4580, 1, [95]], [5021, 1, [2621, [[362, 1, [195]], [390, 10, [18, 21, 76, 206, 209, 238, 263, 342, 362, 374]], [393, 1, [18]], [552, 2, [108, 167]], [560, 1, [80]], [724, 3, [89, 133, 136]], [775, 1, [1753, 1, [4038]], [1959, 1, [574]], [2131, 1, [511]], [2709, 1, [436]], [2792, 4, [18, 182, 223, 302]], [2793, 4, [12, 15, 231, 316]], [4580, 1, [95]], [5021, 1, [2621, [[362, 1, [195]], [390, 10, [18, 21, 76, 206, 209, 238, 263, 342, 362, 374]], [393, 1, [18]], [552, 2, [108, 167]], [560, 1, [80]], [724, 3, [89, 133, 136]], [775, 1, [1753, 1, [4038]], [1959, 1, [574]], [2131, 1, [511]], [2709, 1, [436]], [2792, 4, [18, 182, 223, 302]], [2793, 4, [12, 15, 231, 316]], [4580, 1, [95]], [5021, 1, [2621, [[362, 1, [195]], [390, 10, [18, 21, 76, 206, 209, 238, 263, 342, 362, 374]], [393, 1, [18]], [552, 2, [108, 167]], [560, 1, [80]], [724, 3, [89, 133, 136]], [775, 1, [1753, 1, [4038]], [1959, 1, [574]], [2131, 1, [511]], [2709, 1, [436]], [2792, 4, [18, 182, 223, 302]], [2793, 4, [12, 15, 231, 316]], [4580, 1, [95]], [5021, 1, [2621, [[362, 1, [195]], [390, 10, [18, 21, 76, 206, 209, 238, 263, 342, 362, 374]], [393, 1, [18]], [552, 2, [108, 167]], [560, 1, [80]], [724, 3, [89, 133, 136]], [775, 1, [1753, 1, [4038]], [1959, 1, [574]], [2131, 1, [511]], [2709, 1, [436]], [2792, 4, [18, 182, 223, 302]], [2793, 4, [12, 15, 231, 316]], [4580, 1, [95]], [5021, 1, [2621, [[362, 1, [195]], [390, 10, [18, 21, 76, 206, 209, 238, 263, 342, 362, 374]], [393, 1, [18]], [552, 2, [108, 167]], [560, 1, [80]], [724, 3, [89, 133, 136]], [775, 1, [1753, 1, [4038]], [1959, 1, [574]], [2131, 1, [511]], [2709, 1, [436]], [2792, 4, [18, 182, 223, 302]], [2793, 4, [12, 15, 231, 316]], [4580, 1, [95]], [5021, 1, [2621, [[362, 1, [195]], [390, 10, [18, 21, 76, 206, 209, 238, 263, 342, 362, 374]], [393, 1, [18]], [552, 2, [108, 167]], [560, 1, [80]], [724, 3, [89, 133, 136]], [775, 1, [1753, 1, [4038]], [1959, 1, [574]], [2131, 1, [511]], [2709, 1, [436]], [2792, 4, [18, 182, 223, 302]], [2793, 4, [12, 15, 231, 316]], [4580, 1, [95]], [5021, 1, [2621, [[362, 1, [195]], [390, 10, [18, 21, 76, 206, 209, 238, 263, 342, 362, 374]], [393, 1, [18]], [552, 2, [108, 167]], [560, 1, [80]], [724, 3, [89, 133, 136]], [775, 1, [1753, 1, [4038]], [1959, 1, [574]], [2131, 1, [511]], [2709, 1, [436]], [2792, 4, [18, 182, 223, 302]], [2793, 4, [12, 15, 231, 316]], [4580, 1, [95]], [5021, 1, [2621, [[362, 1, [195]], [390, 10, [18, 21, 76, 206, 209, 238, 263, 342, 362, 374]], [393, 1, [18]], [552, 2, [108, 167]], [560, 1, [80]], [724, 3, [89, 133, 136]], [775, 1, [1753, 1, [4038]], [1959, 1, [574]], [2131, 1, [511]], [2709, 1, [436]], [2792, 4, [18, 182, 223, 302]], [2793, 4, [12, 15, 231, 316]], [4580, 1, [95]], [5021, 1, [2621, [[362, 1, [195]], [390, 10, [18, 21, 76, 206, 209, 238, 263, 342, 362, 374]], [393, 1, [18]], [552, 2, [108, 167]], [560, 1, [80]], [724, 3, [89, 133, 136]], [775, 1, [1753, 1, [4038]], [1959, 1, [574]], [2131, 1, [511]], [2709, 1, [436]], [2792, 4, [18, 182, 223, 302]], [2793, 4, [12, 15, 231, 316]], [4580, 1, [95]], [5021, 1, [2621, [[362, 1, [195]], [390, 10, [18, 21, 76, 206, 209, 238, 263, 342, 362, 374]], [393, 1, [18]], [552, 2, [108, 167]], [560, 1, [80]], [724, 3, [89, 133, 136]], [775, 1, [1753, 1, [4038]], [1959, 1, [574]], [2131, 1, [511]], [2709, 1, [436]], [2792, 4, [18, 182, 223, 302]], [2793, 4, [12, 15, 231, 316]], [4580, 1, [95]], [5021, 1, [2621, [[362, 1, [195]], [390, 10, [18, 21, 76, 206, 209, 238, 263, 342, 362, 374]], [393, 1, [18]], [552, 2, [108, 167]], [560, 1, [80]], [724, 3, [89, 133, 136]], [775, 1, [1753, 1, [4038]], [1959, 1, [574]], [2131, 1, [511]], [2709, 1, [436]], [2792, 4, [18, 182, 223, 302]], [2793, 4, [12, 15, 231, 316]], [4580, 1, [95]], [5021, 1, [2621, [[362, 1, [195]], [390, 10, [18, 21, 76, 206, 209, 238, 263, 342, 362, 374]], [393, 1, [18]], [552, 2, [108, 167]], [560, 1, [80]], [724, 3, [89, 133, 136]], [775, 1, [1753, 1, [4038]], [1959, 1, [574]], [2131, 1, [511]], [2709, 1, [436]], [2792, 4, [18, 182, 223, 302]], [2793, 4, [12, 15, 231, 316]], [4580, 1, [95]], [5021, 1, [2621, [[362, 1, [195]], [390, 10, [18, 21, 76, 206, 209, 238, 263, 342, 362, 374]], [393, 1, [18]], [552, 2, [108, 167]], [560, 1, [80]], [724, 3, [89, 133, 136]], [775, 1, [1753, 1, [4038]], [1959, 1, [574]], [2131, 1, [511]], [2709, 1, [436]], [2792, 4, [18, 182, 223, 302]], [2793, 4, [12, 15, 231, 316]], [4580, 1, [95]], [5021, 1, [2621, [[362, 1, [195]], [390, 10, [18, 21, 76, 206, 209, 238, 263, 342, 362, 374]], [393, 1, [18]], [552, 2, [108, 167]], [560, 1, [80]], [724, 3, [89, 133, 136]], [775, 1, [1753, 1, [4038]], [1959, 1, [574]], [2131, 1, [511]], [2709, 1, [436]], [2792, 4, [18, 182, 223, 302]], [2793, 4, [12, 15, 231, 316]], [4580, 1, [95]], [5021, 1, [2621, [[362, 1, [195]], [390, 10, [18, 21, 76, 206, 209, 238, 263, 342, 362, 374]], [393, 1, [18]], [552, 2, [108, 167]], [560, 1, [80]], [724, 3, [89, 133, 136]], [775, 1, [1753, 1, [4038]], [1959, 1, [574]], [2131, 1, [511]], [2709, 1, [436]], [2792, 4, [18, 182, 223, 302]], [2793, 4, [12, 15, 231, 316]], [4580, 1, [95]], [5021, 1, [2621, [[362, 1, [195]], [390, 10, [18, 21, 76, 206, 209, 238, 263, 342, 362, 374]], [393, 1, [18]], [552, 2, [108, 167]], [560, 1, [80]], [724, 3, [89, 133, 136]], [775, 1, [1753, 1, [4038]], [1959, 1, [574]], [2131, 1, [511]], [2709, 1, [436]], [2792, 4, [18, 182, 223, 302]], [2793, 4, [12, 15, 231, 316]], [4580, 1, [95]], [5021, 1, [2621, [[362, 1, [195]], [390, 10, [18, 21, 76, 206, 209, 238, 263, 342, 362, 374]], [393, 1, [18]], [552, 2, [108, 167]], [560, 1, [80]], [724, 3, [89, 133, 136]], [775, 1, [1753, 1, [4038]], [1959, 1, [574]], [2131, 1, [511]], [2709, 1, [436]], [2792, 4, [18, 182, 223, 302]], [2793, 4, [12, 15, 231, 316]], [4580, 1, [95]], [5021, 1, [2621, [[362, 1, [195]], [390, 10, [18, 21, 76, 206, 209, 238, 263, 342, 362, 374]], [393, 1, [18]], [552, 2, [108, 167]], [560, 1, [80]], [724, 3, [89, 133, 136]], [775, 1, [1753, 1, [4038]], [1959, 1, [574]], [2131, 1, [511]], [2709, 1, [436]], [2792, 4, [18, 182, 223, 302]], [2793, 4, [12, 15, 231, 316]], [4580, 1, [95]], [5021, 1, [2621, [[362, 1, [195]], [390, 10, [18, 21, 76, 206, 209, 238, 263, 342, 362, 374]], [393, 1, [18]], [552, 2, [108, 167]], [560, 1, [80]], [724, 3, [89, 133, 136]], [775, 1, [1753, 1, [4038]], [1959, 1, [574]], [2131, 1, [511]], [2709, 1, [436]], [2792, 4, [18, 182, 223, 302]], [2793, 4, [12, 15, 231, 316]], [4580, 1, [95]], [5021, 1, [2621, [[362, 1, [195]], [390, 10, [18, 21, 76, 206, 209, 238, 263, 342, 362, 374]], [393, 1, [18]], [552, 2, [108, 167]], [560, 1, [80]], [724, 3, [89, 133, 136]], [775, 1, [1753, 1, [4038]], [1959, 1, [574]], [2131, 1, [511]], [2709, 1, [436]], [2792, 4, [18, 182, 223, 302]], [2793, 4, [12, 15, 231, 316]], [4580, 1, [95]], [5021, 1, [2621, [[362, 1, [195]], [390, 10, [18, 21, 76, 206, 209, 238, 263, 342, 362, 374]], [393, 1, [18]], [552, 2, [108, 167]], [560, 1, [80]], [724, 3, [89, 133, 136]], [775, 1, [1753, 1, [4038]], [1959, 1, [574]], [2131, 1, [511]], [2709, 1, [436]], [2792, 4, [18, 182, 223, 302]], [2793, 4, [12, 15, 231, 316]], [4580, 1, [95]], [5021, 1, [2621, [[362, 1, [195]], [390, 10, [18, 21, 76, 206, 209, 238, 263, 342, 362, 374]], [393, 1, [18]], [552, 2, [108, 167]], [560, 1, [80]], [724, 3, [89, 133, 136]], [775, 1, [1753, 1, [4038]], [1959, 1, [574]], [2131, 1, [511]], [2709, 1, [436]], [2792, 4, [18, 182, 223, 302]], [2793, 4, [12, 15, 231, 316]], [4580, 1, [95]], [5021, 1, [2621, [[362, 1, [195]], [390, 10, [18, 21, 76, 206, 209, 238, 263, 342, 362, 374]], [393, 1, [18]], [552, 2, [108, 167]], [560, 1, [80]], [724, 3, [89, 133, 136]], [775, 1, [1753, 1, [4038]], [1959, 1, [574]], [2131, 1, [511]], [2709, 1, [436]], [2792, 4, [18, 182, 223, 302]], [2793, 4, [12, 15, 231, 316]], [4580, 1, [95]], [5021, 1, [2621, [[362, 1, [195]], [390, 10, [18, 21, 76, 206, 209, 238, 263, 342, 362, 374]], [393, 1, [18]], [552, 2, [108, 167]], [560, 1, [80]], [724, 3, [89, 133, 136]], [775, 1, [1753, 1, [4038]], [1959, 1, [574]], [2131, 1, [511]], [2709, 1, [436]], [2792, 4, [18, 182, 223, 302]], [2793, 4, [12, 15, 231, 316]], [4580, 1, [95]], [5021, 1, [2621, [[362, 1, [195]], [390, 10, [18, 21, 76, 206, 209, 238, 263, 342, 362, 374]], [393, 1, [18]], [552, 2, [108, 167]], [560, 1, [80]], [724, 3, [89, 133, 136]], [775, 1, [1753, 1, [4038]], [1959, 1, [574]], [2131, 1, [511]], [2709, 1, [436]], [2
```

(حالت ت)

The screenshot shows a JupyterLab environment with a dark theme. At the top, the file browser shows 'IR_phase1.ipynb' with a star icon. The main area contains a terminal window with the following content:

```

enter a word for checking: واکن استراژیکا
Normal word to search استراژیکا
594 results in 50.772 ms
id -> title

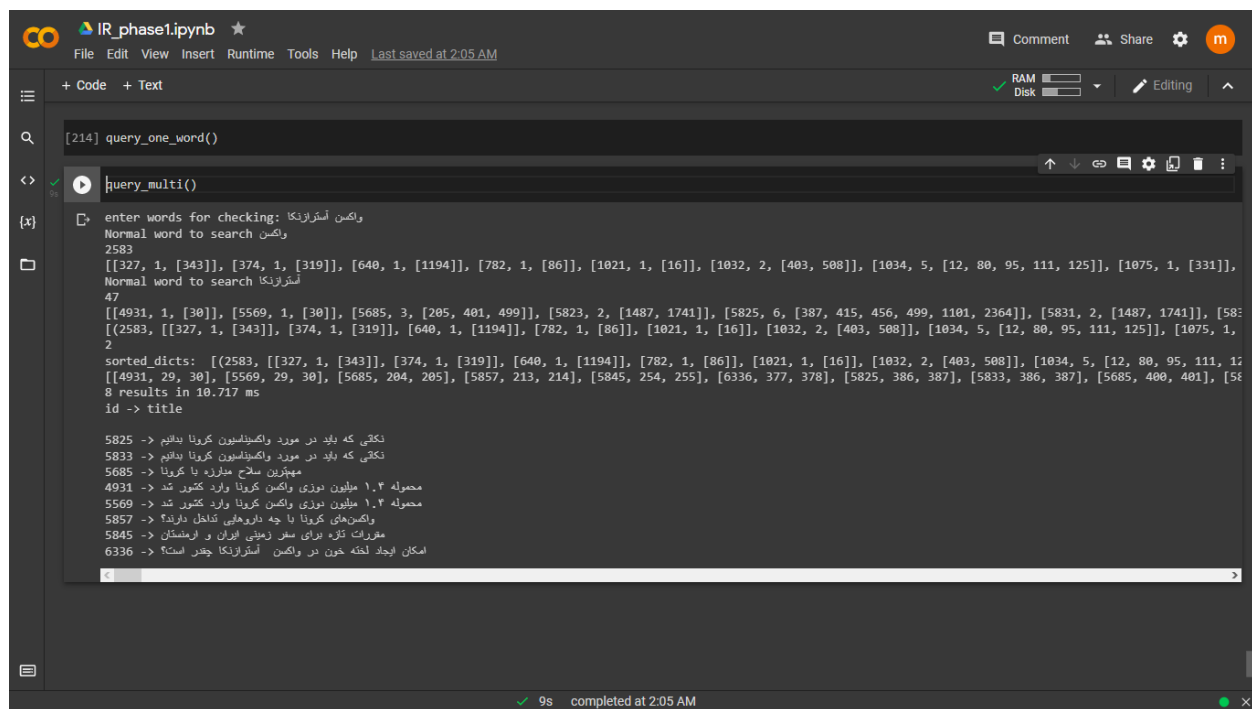
The Five Top Related Docs

5889 -> آمار تفکیکی واکنسیناسیون کرونا در ایران تا ۲۶ مردامه
5896 -> آمار تفکیکی واکنسیناسیون کرونا در ایران تا ۲۶ مردامه
5936 -> اعلام آمار تفکیکی واکنسیناسیون کرونا تا صبح امروز
5864 -> آمار تفکیکی واکنسیناسیون کرونا در کشور تا ۲۱ مرداد
5853 -> آمار تفکیکی واکنسیناسیون کرونا در کشور تا ۲۱ مرداد
به گزارش ایسا دکتر کیلوش جهانبور رئیس مرکز روابط عمومی و اطلاع رسانی وزارت بهداشت آمار تفکیکی واکنسیناسیون کرونا در کشور از ابتدای برنامه ملی واکنسیناسیون علیه کوید تا مرداد را اعلام کرد
سنوفاوم نوبت اول
سنوفاوم نوبت دوم
استراژیکا نوبت اول
استراژیکا نوبت دوم
برکت نوبت اول
برکت نوبت دوم
اسپنتیک نوبت اول
اسپنتیک نوبت دوم
بهرات نوبت اول
بهرات نوبت دوم
آمار تفکیکی واکنسیناسیون کرونا در ایران تا ۲۶ مردامه
به گزارش ایسا دکتر کیلوش جهانبور رئیس مرکز روابط عمومی و اطلاع رسانی وزارت بهداشت آمار تفکیکی واکنسیناسیون کرونا در کشور از ابتدای برنامه ملی واکنسیناسیون علیه کوید تا مرداد را اعلام کرد
سنوفاوم نوبت اول
سنوفاوم نوبت دوم
استراژیکا نوبت اول
استراژیکا نوبت دوم

```

The status bar at the bottom indicates '3s completed at 1:47 AM'.

فاز ۱



The screenshot shows a Jupyter Notebook with a file named 'IR_phaset1.ipynb'. The code cell contains a call to 'query_multi()' which returns a list of results. The results include a list of words and their corresponding IDs, a list of sorted dictionaries, and a list of results in Persian. The Persian results are as follows:

- 5825 -> نکاتی که باید در مورد واکسیناسیون کرونا بدانیم
- 5833 -> نکاتی که باید در مورد واکسیناسیون کرونا بدانیم
- 5685 -> بهترین سلاح مبارزه با کرونا
- 4931 -> محموله ۱.۴ میلیون دوزی واکسن کرونا وارد کشور شد
- 5569 -> محموله ۱.۴ میلیون دوزی واکسن کرونا وارد کشور شد
- 5857 -> واکسن‌های کرونا با چه داروهایی تداخل دارند؟
- 5845 -> مقررات تازه برای سفر زمینی ایران و ارمنستان
- 6336 -> امکان ایجاد لخته خون در واکسن آسترانکا چقدر است؟

هر دو سند مرتبط هستند اما نتیجه در حالت ۲ مرتبط تر است در فاز ۱ بیشتر به بودن دو کلمه توجه شده در حالی که در حالت دوم به این توجه شده است که کلمه آسترانکا با توجه به این که خاص تر هست *idf* بیشتری دارد و تاثیر زیادی دارد و خبر در مورد نوبت دهی این واکسن آورده شده است و رتبه بالاتر نسبت به قبلی دارد.

کلمات دیگر که در فاز یک آمده بود نیز بررسی شدن که به همانگونه برای رتبه دهی بیشتر به مرتبط بودن آن سند دارد

IR phasel.pyynb

File Edit View Insert Runtime Tools Help

Comment Share

RAM Disk

Editing

+ Code + Text

```
time_finish = datetime.datetime.now()
print("{} results in {}".format(len(Scores), ((time_finish - time_start).total_seconds()*1000))
Show_res(Scores, K, normal_query)
```

enter a word for checking: جمهوری اسلامی ایران

Normal word to search ایران

4004 results in 2298.5319999999997 ms

id -> title

The Five Top Related Docs

رئیس جمهور یک قانون مجلس را برای اجرا به وزارت نفت و سازمان برنامه و بودجه ابلاغ کرد -> 2479

رئیس جمهور قانون جهش تولید مسکن را برای اجرا ابلاغ کرد -> 2736

آیت الله رئیسی یک قانون مجلس را برای اجرا ابلاغ کرد -> 2589

مجلس جدید جمهوری آذربایجان در ایران استوانه خود را تقدیم رئیس‌جمهور کرد -> 1701

امضای استاد همکاری دو جاقبه بین مقامات ایران و تاجیکستان -> 2904

مجلس شورای اسلامی را برای اجرا به وزارت نفت و سازمان برنامه و بودجه کشور ابلاغ کرد

رئیس جمهور یک قانون مجلس را برای اجرا به وزارت نفت و سازمان برنامه و بودجه ابلاغ کرد

به گزارش حوزه دولت خبرگزاری فارس آیت‌الله سید ابراهیم رئیسی سید ابراهیم رئیسی رئیس جمهور در اجرای اصل قانون اساسی جمهوری اسلامی ایران قانون تفسیر بند ز تبصره قانون بودجه سال گل کشور مصوب مجلس شورای اسلامی مجلس شورای اسلامی را برای اجرا به وزارت نفت و سازمان برنامه و بودجه کشور ابلاغ کرد

[214] query_one_word()

[223] query_multi()

11s

completed at 2:12 AM

IR phasel.pyynb

File Edit View Insert Runtime Tools Help All changes saved

Comment Share

RAM Disk

Editing

+ Code + Text

```
time_finish = datetime.datetime.now()
print("{} results in {}".format(len(Scores), ((time_finish - time_start).total_seconds()*1000))
Show_res(Scores, K, normal_query)
```

enter a word for checking: سازمان ملل متحد

Normal word to search سازمان ملل متحد

129 results in 9.539000000000001 ms

id -> title

The Five Top Related Docs

هیئت پارلمانی ایران به رم سفر کرد -> 2301

رایزنی وزیران امور خارجه ایران و واترلاند فردا در تهران -> 4772

گزارش نظارت میدانی نمایندگان از مرزهای شمال‌غرب به کمیسیون امنیت ملی -> 1973

شهرهای میوبای -> 6799

انتقال سکونتگاه و حقوق بشر» برای بیهوا -> 6911

به گزارش حوزه پارلمانی خبرگزاری فارس هیئت پارلمانی جمهوری اسلامی ایران به منظور شرکت در نشست مقدماتی کاپ کنفرانس کنفرانس آب و هوا و اقلیمی به شهر رم سفر کرد

نشست مقدماتی کاپ با حضور رؤسا و نمایندگان پارلمان‌های کشور و سازمان بین‌المللی به میزبانی مجلس نمایندگان ایتالیا و با مشارکت اتحادیه بین‌المجلس جهانی و سازمان ملل متحد در روزهای و مهر ماه برگزار خواهد شد

این کنفرانس در خصوص اهداف جهانی جهت مقابله با تغییرات اقلیمی رویکرد سبز جهت بهبود وضعیت زیست‌ی در دوران کووید ۲۰۲۰ تأمین مالی سیاست‌های جهانی در حوزه تغییر اقلیم و نقش پارلمان‌ها در دستیابی به اهداف کاپ گشتگو خواهد شد

قراردست پیش‌پایان اولیه سند نهایی کنفرانس پارلمانی سازمان ملل متحد در خصوص تغییرات اقلیم در این کنفرانس به تصویب رسید

[214] query_one_word()

[223] query_multi()

7s

completed at 2:12 AM

```
IR_phasel.ipynb
File Edit View Insert Runtime Tools Help All changes saved

+ Code + Text
normal_query = lstemmer.stem(query)
print("Normal word to search {}".format(normal_query))
query_dict = vectorize_query(query)
K = 5
Scores = CosineScore(query_dict, dictionarys, doc_norm, K)
time_finish = datetime.datetime.now()
print("{} results in {} ms".format(len(Scores), ((time_finish - time_start).total_seconds()*1000))
Show_res(Scores, K, normal_query)

دانشگاه صنعتی امیرکبیر
دانشگاه صنعتی امیرکبیر
Normal word to search امیرکبیر
933 results in 111.86500000000001 ms
id -> title

The Five Top Related Docs

2703 -> باید برای ثبت نقش دانشگاهیان در دوران دفاع مقدس کار تحقیقی صورت گیرد ->
2702 -> امروز محیط دانشگاهی ما عرصه دفاع مقدس است ->
دفترچه راهنمای آزمون استخدامی دانشگاهها برای بار چهارم اصلاح شد/تجدید مهلت ثبتنام ->
دفترچه راهنمای آزمون استخدامی دانشگاهها برای بار چهارم اصلاح شد/تجدید مهلت ثبتنام ->
اصطلاح گرامی درجه یک هری به بسند بدلایی ->
7520 ->
سورج خلی مهم هست که بدایم دانشگاهیان دانشگاهیان در تاریخ انقلاب و دوران دفاع مقدس چه نقشی داشتند و جای از آن وظیفه اصلی که از آنها انتظار میرفت و تحمیل در دانشگاه است چه اقداماتی برای جامعه و انقلاب داشتند
است و ما همواره در این مسئله تاکید داریم ولی در کنار این موضوع دانشجویی که فارغ التحصیل میشوند هر چند با مسائل اجتماعی سیاسی فرهنگی جامعه اشتغال دارند میفرماد نقش موثرتری را در کنار کارهای علمی در جامعه ایفا کند
که با حکومت طاغوت مبارزه کردند از همین دانشگاهها بودند دانشجویان و جوانان در دوران انقلاب نقش و نژای داشتند بسیاری از قیامها و مستندات را که نگاه میکنیم میبینیم اکثر جوانان نقش داشتند حفظ انقلاب هم در دوش جوانان است
در ادامه با اشاره به نقش دانشجویان و جوانان در تشکیل نیروهای بسیج بسیج و میانه گشت و حتی خاطرات فرماندهان بزرگ را مشاهده میکنیم میبینیم که آنها دانشجویانی بودند که به جبهه های دفاع مقدس رفتند و در مقابل دشمن ایستادند
متممیت افزود اگر بخواهیم در مورد نقش دانشجو و دانشگاه در جبهه های جنگ و در دوران دفاع مقدس صحبت کنیم نواز به یک کار تحقیقی است که توصیه میکنم به این کار پرداخته شود
وی افزود پیش از شهید دوران دفاع مقدس از دانشگاه امیرکبیر بودند یعنی در زمانی که دانشگاه حدود دانشجو داشت دانشگاهیان این دانشگاه این گروه در خدمت جبهه و جنگ بودند
در دوران دفاع مقدس جهت انتقال به نسل جدید بسیار اهمیت دارد البته در بخش های مختلف این اتفاق افتاده و اگر یک تیم محقق بشنید و خدمت دانشگاهیان در زمان جنگ تحمیلی را ثبت کنند کار بسیار خوبی خوبی انجام خواهد شد
باید برای ثبت نقش دانشگاهیان در دوران دفاع مقدس کار تحقیقی صورت گیرد
به گزارش خبرنگار تسکیمهای دانشگاهی خبرنگاری فارس سیاحمد معتمدی رئیس دانشگاه
صنعتی امیرکبیر
دانشگاه صنعتی امیرکبیر
نقشهایی که به همت نهاد نمایندگی مقام معظم رهبری در دانشگاهها به مناسبت فرارسیدن
هفته دفاع مقدس برگزار شد درباره نقش آموزشی علمی در حفظ و نگهداری خاطرات دوران
دفاع مقدس و یاد و خاطره شهادت جنگ تحمیلی گفت این موضوع خلی مهم هست که بدایم
دانشگاهیان دانشگاهیان در تاریخ انقلاب و
دوران دفاع مقدس چه نقشی داشتند و جای از آن وظیفه اصلی که از آنها انتظار
```

10s completed at 2:12 AM

که با توجه به *content* ها میتوان تشخیص داد واکشی ها بهتر و مرتبط تر شده است در کل.

بخش امتیازی

در این بخش ابتدا ما دیتاهای خود را مورد پردازش قرار می‌دهیم به این صورت که تمامی کلمات رو می‌دهیم تا **word vector** آنها رو بدست بیاریم حال برای وزن دهی آنها از همان مقادیر *tfidf* خودمان که در همین فاز بدست آورده‌ایم استفاده می‌کنیم.

زمانی که می‌خواهیم *doc embbeding* رو بدست بیاریم می‌توانیم کلماتی که موجود نیست رو در نظر نگیریم (با *try except*)

حالا برای چند کوئری می‌ایم ببینم مدلی که آموزش دادیم چه کرد

ژیمناستیک

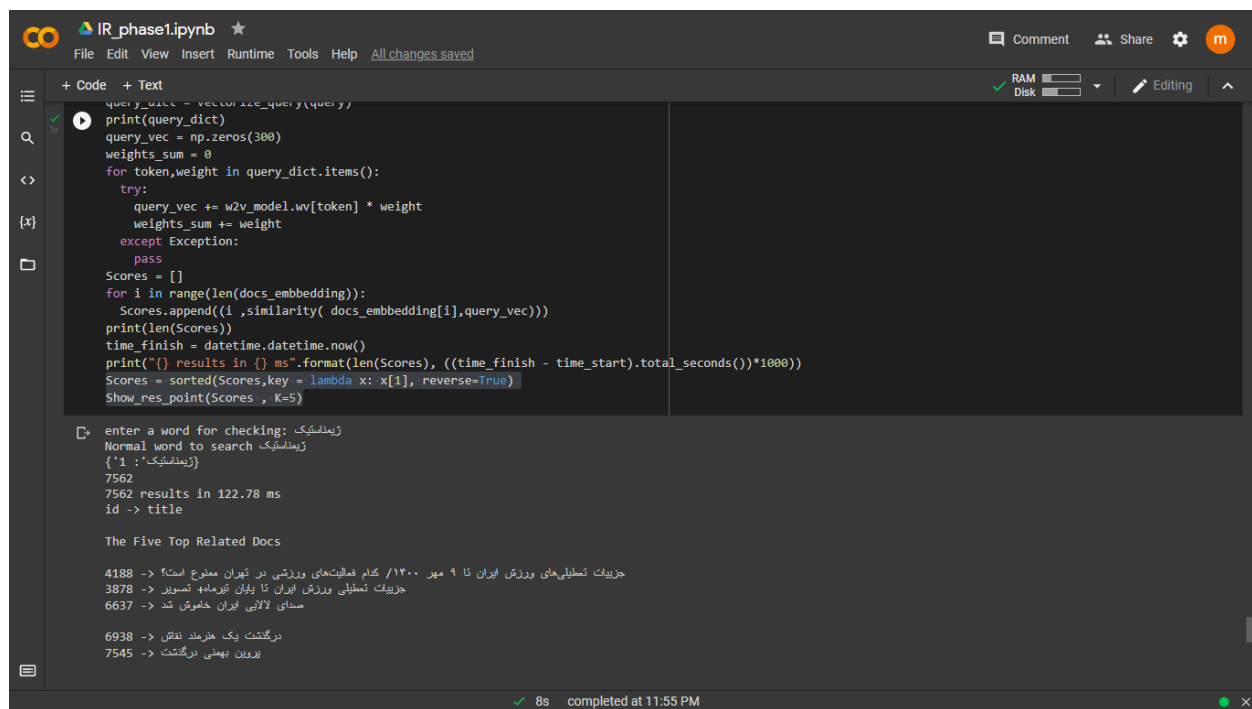
در ۵ تا داک باز گردانده شده در مدلی که ما آموزش دادیم به این صورت هست که مرتبط ترین داک در مورد تعطیلی رشته‌های ورزشی هست و سند بعدی هم به همین صورت ولی سند سوم به فوت هنرمند پیشکسوت بوده که اصلا مرتبط نبوده (تنها شباهتی که دارد وجود هنر بوده 😊) و دو داک بعدی هم به همین صورت بوده

برای

(۱) میشود برای این حالت زیرا اولینی که بازگشته مرتبط هست) $Precision@K = 1$

(۰/۴) می‌شود برای این حالت زیرا که فقط ۲ تا از ۵ تا مرتبط بوده) $Precision@K = 5$

در اینجا دو تا اولی فقط مرتبط بود و مقدار ۰/۷۵ درصد می‌شود $MAP (0.5(1+1/2))$



```
query_vec = vectorize_query(query)
print(query_dict)
query_vec = np.zeros(300)
weights_sum = 0
for token, weight in query_dict.items():
    try:
        query_vec += w2v_model.wv[token] * weight
        weights_sum += weight
    except Exception:
        pass
Scores = []
for i in range(len(docs_embedding)):
    Scores.append((i, similarity(docs_embedding[i], query_vec)))
print(len(Scores))
time_finish = datetime.datetime.now()
print("{} results in {} ms".format(len(Scores), ((time_finish - time_start).total_seconds()*1000))
Scores = sorted(Scores, key = lambda x: x[1], reverse=True)
Show_res_point(Scores, K=5)
```

enter a word for checking: زمینشیک
Normal word to search زمینشیک
{'1': 'زمینشیک'}
7562
7562 results in 122.78 ms
id -> title

The Five Top Related Docs

4188 جزئیات تعطیلاتی ورزش ایران تا ۹ مهر ۱۴۰۰ / کدام تعطیلاتی ورزشی در تهران ممنوع است؟ ->
3878 جزئیات تعطیلاتی ورزش ایران تا پایان تیرماه تصویر ->
6637 صدای لایلی ایران خاموش شد ->
6938 درگشت یک هزارمده نقش ->
7545 یروین بهمنی درگشت ->

8s completed at 11:55 PM

در ۵ داک اول که با مدل از پیش آموزش داده استفاده کردیم به این صورت هست که هر ۵ داک به صورتی با کوئری مد نظر ما ارتباط داشتند و هیچ داک بی ارتباط نبود

(۱) میشود برای این حالت زیر اولینی که بازگشته مرتبط هست) $Precision@K = 1$

(هر ۵ داک مرتبط اند پس ۱ می شود) $Precision@K = 5$

هم ۱ می شود $MAP(0.2(1+1+1+1+1))$

The screenshot shows a Jupyter Notebook with a file named 'IR_phasel.ipynb'. The code in the notebook performs a word search and displays related documents. The output shows the search results for the word 'ژیمناستیک' (Gymnastics).

```
print('normal word to search {}'.format(normal_query))
query_dict = vectorize_query(query)
print(query_dict)
query_vec = np.zeros([300])
weights_sum = 0
for token, weight in query_dict.items():
    try:
        query_vec += w2v_model_150K.wv[token] * weight
        weights_sum += weight
    except Exception:
        pass
Scores = []
for i in range(len(docs_embedding_150K)):
    Scores.append((i, similarity(docs_embedding_150K[i], query_vec)))
print(len(Scores))
time_finish = datetime.datetime.now()
print("{} results in {} ms".format(len(Scores), ((time_finish - time_start).total_seconds())*1000))
Scores = sorted(Scores, key = lambda x: x[1], reverse=True)
Show_res_point(Scores, K=5)
```

enter a word for checking: ژیمناستیک
Normal word to search ژیمناستیک
{'1': 'ژیمناستیک'}
7562
7562 results in 112.271 ms
id -> title

The Five Top Related Docs

- 3484 -> رشته ورزشی زور تبح پروتکل‌های کرونا ۲۳
- 3878 -> جزئیات تسلیلی ورزش ایران تا پایان نبردها تصویر ->
- 4188 -> جزئیات تسلیلی‌های ورزش ایران تا ۹ مهر ۱۴۰۰ کدام فعالیت‌های ورزشی در تهران ممنوع است؟ ->
- 3615 -> دبیر مجمع فراسیون ژیمناستیک مشخص شد ->
- 632 -> خیرخواه: برخی به دنبال فلج کردن ژیمناستیک هستند/ با باکیوت فراسیون موفقیت‌ها بیشتر شد ->

در مورد کوئری بعد واکسن آسترازنکا در نظر گرفتیم

در هر دو مدل ۵ داک ابتدایی بازگردانده شده یکسان هست ولی تفاوت در رنگ بندی موجود هست و این می‌تواند به خاطر خاص بودن کوئری ما بوده

در این حالت مدل که ما آموزش دادیم

```

query_dict = vectorize_query(query)
print(query_dict)
query_vec = np.zeros(300)
weights_sum = 0
for token, weight in query_dict.items():
    try:
        query_vec += w2v_model.wv[token] * weight
        weights_sum += weight
    except Exception:
        pass
Scores = []
for i in range(len(docs_embedding)):
    Scores.append((i, similarity(docs_embedding[i], query_vec)))
print(len(Scores))
time_finish = datetime.datetime.now()
print("{} results in {} ms".format(len(Scores), ((time_finish - time_start).total_seconds()*1000))
Scores = sorted(Scores, key = lambda x: x[1], reverse=True)
Show_res_point(Scores, K=5)

```

enter a word for checking: واکن
 Normal word to search واکن
 {'1': 'واکن', '1': 'واکن'}
 7562
 7562 results in 100.29100000000001 ms
 id -> title
 The Five Top Related Docs
 5913 -> آمار تفکیکی واکنسیون کرونا در کشور تا ۲ شهریور
 5821 -> آمار تولید، تحویل و توزیع واکن کووایران برکت تا امروز
 5828 -> آمار تولید، تحویل و توزیع واکن کووایران برکت تا امروز
 5970 -> توزیع بیش از ۳۱ میلیون کُر واکن کرونا در کشور تا کنون
 5955 -> اعلام آمار تفکیکی واکنسیون کرونا در کشور تا ۱۴ شهریور ماه

و مدل از پیش آموزش داده شده

```

print(normal_word_to_search).format(normal_query))
query_dict = vectorize_query(query)
print(query_dict)
query_vec = np.zeros(300)
weights_sum = 0
for token, weight in query_dict.items():
    try:
        query_vec += w2v_model_150K.wv[token] * weight
        weights_sum += weight
    except Exception:
        pass
Scores = []
for i in range(len(docs_embedding_150K)):
    Scores.append((i, similarity(docs_embedding_150K[i], query_vec)))
print(len(Scores))
time_finish = datetime.datetime.now()
print("{} results in {} ms".format(len(Scores), ((time_finish - time_start).total_seconds()*1000))
Scores = sorted(Scores, key = lambda x: x[1], reverse=True)
Show_res_point(Scores, K=5)

```

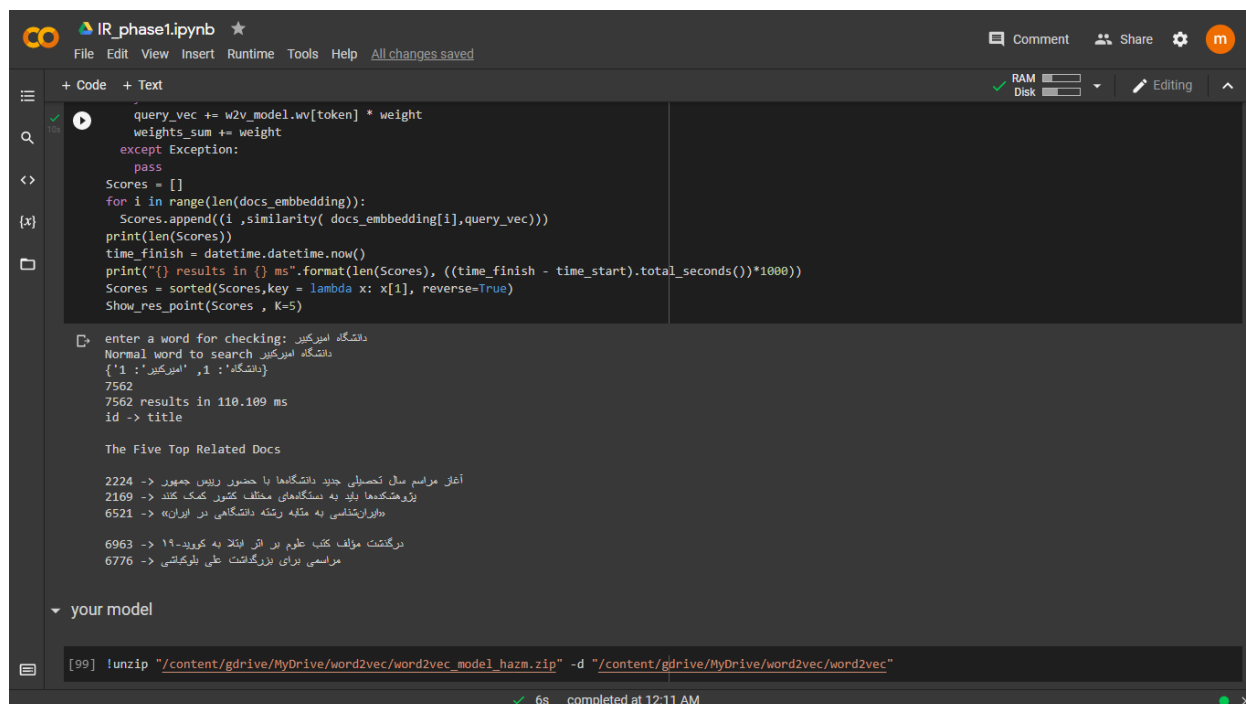
enter a word for checking: واکن
 Normal word to search واکن
 {'1': 'واکن', '1': 'واکن'}
 7562
 7562 results in 92.744 ms
 id -> title
 The Five Top Related Docs
 5913 -> آمار تفکیکی واکنسیون کرونا در کشور تا ۲ شهریور
 5970 -> توزیع بیش از ۳۱ میلیون کُر واکن کرونا در کشور تا کنون
 5821 -> آمار تولید، تحویل و توزیع واکن کووایران برکت تا امروز
 5828 -> آمار تولید، تحویل و توزیع واکن کووایران برکت تا امروز
 5955 -> اعلام آمار تفکیکی واکنسیون کرونا در کشور تا ۱۴ شهریور ماه

۱) میشود برای این حالت زیرا اولینی که بازگشته مرتبط هست) $Precision@K = 1$

(هر ۵ داک مرتبط اند پس ۱ می‌شود) $Precision@K = 5$

هم ۱ می‌شود $MAP(0.2(1+1+1+1+1))$

برای دانشگاه امیرکبیر که کمی عام تر از قبلی هست به این صورت هست
در مدلی که ما آزمایش انجام دادیم



```
query_vec += w2v_model.wv[token] * weight
weights_sum += weight
except Exception:
    pass
Scores = []
for i in range(len(docs_embedding)):
    Scores.append((i, similarity(docs_embedding[i], query_vec)))
print(len(Scores))
time_finish = datetime.datetime.now()
print("{} results in {} ms".format(len(Scores), ((time_finish - time_start).total_seconds()*1000))
Scores = sorted(Scores, key = lambda x: x[1], reverse=True)
Show_res_point(Scores , K=5)
```

دانشگاه امیرکبیر
دانشگاه امیرکبیر
{'1': 'امیرکبیر', 1: 'دانشگاه'}
7562
7562 results in 110.109 ms
id -> title

The Five Top Related Docs

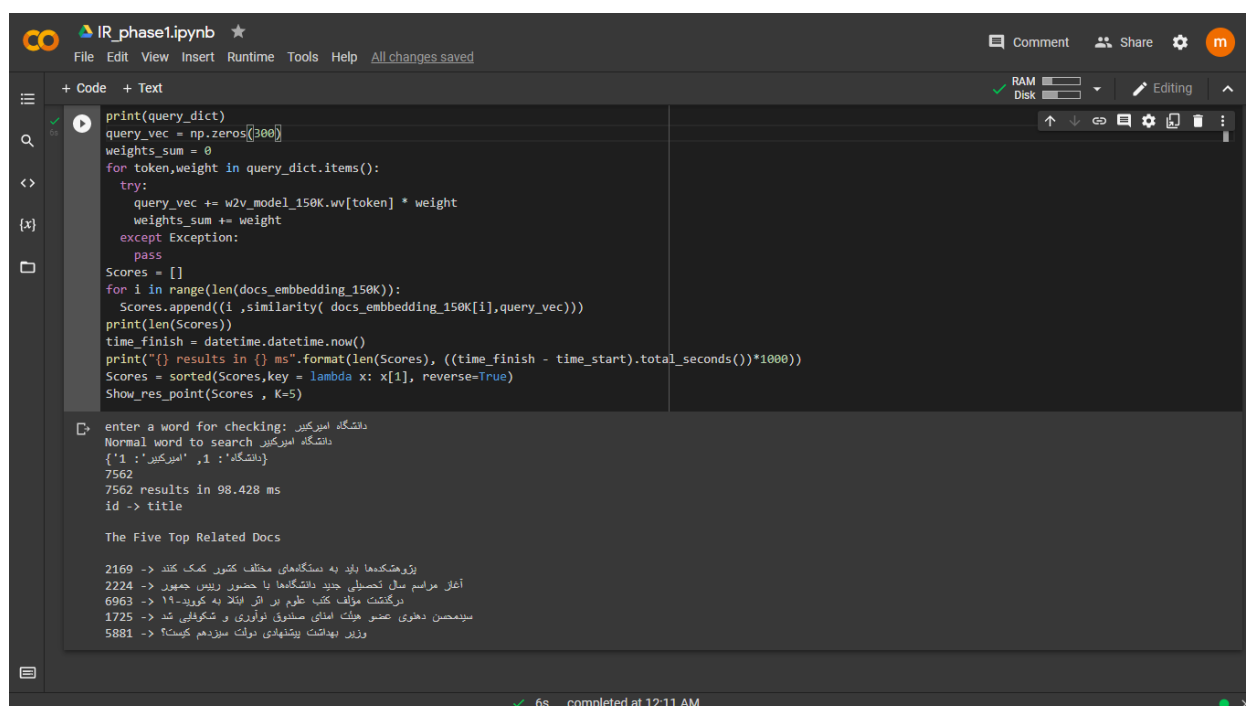
آغاز مراسم سال تحصیلی جدید دانشگاهها با حضور رئیس جمهور -> 2224
پژوهشگاهها باید به دستگاههای مختلف کشور کمک کنند -> 2169
«ایران‌شناسی به مثابه رشته دانشگاهی در ایران» -> 6521
درگشت مؤلف کتب علوم در اثر ابتلا به کرونا-۱۹ -> 6963
مواسمی برای بزرگداشت علی باباگنجی -> 6776

your model

```
[99]: !unzip "/content/gdrive/MyDrive/word2vec/word2vec_model_hazm.zip" -d "/content/gdrive/MyDrive/word2vec/word2vec"
```

6s completed at 12:11 AM

و در مدل آموزش شده



```
print(query_dict)
query_vec = np.zeros([300])
weights_sum = 0
for token, weight in query_dict.items():
    try:
        query_vec += w2v_model_150K.wv[token] * weight
        weights_sum += weight
    except Exception:
        pass
Scores = []
for i in range(len(docs_embedding_150K)):
    Scores.append((i, similarity(docs_embedding_150K[i], query_vec)))
print(len(Scores))
time_finish = datetime.datetime.now()
print("{} results in {} ms".format(len(Scores), ((time_finish - time_start).total_seconds()*1000))
Scores = sorted(Scores, key = lambda x: x[1], reverse=True)
Show_res_point(Scores , K=5)
```

دانشگاه امیرکبیر
دانشگاه امیرکبیر
{'1': 'امیرکبیر', 1: 'دانشگاه'}
7562
7562 results in 98.428 ms
id -> title

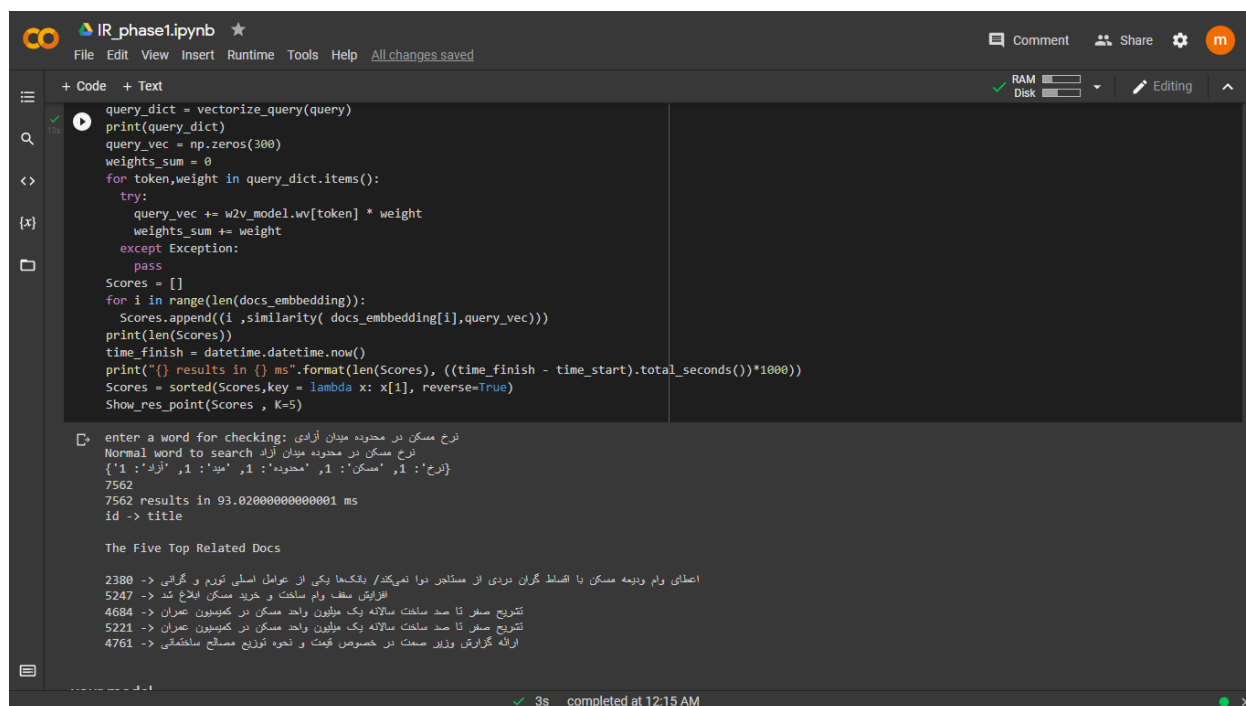
The Five Top Related Docs

پژوهشگاهها باید به دستگاههای مختلف کشور کمک کنند -> 2169
آغاز مراسم سال تحصیلی جدید دانشگاهها با حضور رئیس جمهور -> 2224
درگشت مؤلف کتب علوم در اثر ابتلا به کرونا-۱۹ -> 6963
سیمنصحن دهنوی عضو هیات امنای صندوق نوآوری و شکوفایی شد -> 1725
وزیر بهداشت پیشنهادی دولت سیزدهم چیست؟ -> 5881

6s completed at 12:11 AM

برای نرخ مسکن در محدوده‌ی میدان آزادی

در مدلی که ما آموزش دادیم مهمترین سند واکنشی شده اعطای وام ودیعه مسکن می‌باشد که به صورت کلی مربوط هست ولی نیاز ما رو برطرف نمی‌کند در دیگر حالت ها هم همین گونه است.



```
query_dict = vectorize_query(query)
print(query_dict)
query_vec = np.zeros(300)
weights_sum = 0
for token, weight in query_dict.items():
    try:
        query_vec += w2v_model.wv[token] * weight
        weights_sum += weight
    except Exception:
        pass
Scores = []
for i in range(len(docs_embedding)):
    Scores.append((i, similarity(docs_embedding[i], query_vec)))
print(len(Scores))
time_finish = datetime.datetime.now()
print("{} results in {} ms".format(len(Scores), ((time_finish - time_start).total_seconds()*1000))
Scores = sorted(Scores, key = lambda x: x[1], reverse=True)
Show_res_point(Scores, K=5)
```

enter a word for checking: نرخ مسکن در محدوده میدان آزادی
Normal word to search: نرخ مسکن در محدوده میدان آزاد
{'1': 'نرخ', '1': 'مسکن', '1': 'محدوده', '1': 'میدان', '1': 'آزاد'}
7562
7562 results in 93.02000000000001 ms
id -> title

The Five Top Related Docs

2380 -> اعطای وام ودیعه مسکن با هسلط گران دودی از مستاجر دوا نمی‌گذرد / بانکها یکی از عوامل اصلی تورم و گرانی ->
5247 -> افزایش سقف وام ساخت و خرید مسکن ابلاغ شد ->
4684 -> تئوری سفر تا صد ساخت سالانه یک میلیون واحد مسکن در کمیسیون عمران ->
5221 -> تئوری سفر تا صد ساخت سالانه یک میلیون واحد مسکن در کمیسیون عمران ->
4761 -> ارائه گزارش وزیر صنعت در خصوص قیمت و نحوه توزیع مسکنهای ساختمانی ->

(اولین داک باز گردانده شده کاملاً مرتبط نیست) $Precision@K = 0$

(هر ۵ داک غیر مرتبط اند پس ۰ می‌شود) $Precision@K = 0$

هم ۰ می‌شود $MAP(0, 0)$

در مدل از قبل آموزش شده خیلی به صورت قبلی هست با این فرق که سند دوم و سوم در مورد مسکن هم نیست و در مورد قیمت طلا و سکه است

The screenshot shows a Jupyter Notebook with a file named 'IR_phasel.ipynb'. The code in the notebook performs the following steps:

- Prints the 'normal word to search' (which is 'آزادی').
- Creates a dictionary for the query ('query_dict') and a vector ('query_vec').
- Iterates through the items of 'query_dict' to calculate weights for the query vector.
- Calculates the similarity between the query vector and the document embeddings ('docs_embedding_150K').
- Prints the number of results (7562) and the time taken (118.202 ms).
- Sorts the results by score and prints the top 5 related documents.

The output of the notebook shows the following results:

```

enter a word for checking: آزادی
Normal word to search آزادی
{'1': 'آزاد', '1': 'مید', '1': 'محدوده', '1': 'مسکن'}
7562
7562 results in 118.202 ms
id -> title

The Five Top Related Docs

5018 تسهیلات طرح اقدام ملی مسکن به ۲۵۰ میلیون تومان رسید ->
5263 آفت و خن ملحد قومت سکه و طلا / نوسان ۱۰ درصد خدب سکه ->
5193 افزایش ۱۵۰ هزار تومانی قیمت سکه پس از تسهیلات گروایی/ ویشیایی قیمت‌ها طی روزهای آتی ->
4761 ارائه گزارش وزیر سعت در خصوص قیمت و نحوه توزیع مصالح ساختمانی ->
5387 ارائه گزارش وزیر سعت در خصوص قیمت و نحوه توزیع مصالح ساختمانی ->
  
```

(اولین داک باز گردانده شده کاملاً مرتبط نیست) $Precision@K = ۰$

(هر ۵ داک غیر مرتبط اند پس ۰ می‌شود) $Precision@K = ۰$

هم ۰ می‌شود $MAP(۰, ۰)$

اگر مرتبط بودن رو همین در نظر بگیریم که در مورد مسکن صحبت شده مقادیر متفاوت اند.

برای کوئری پرسشی از این کوئری استفاده کردیم
 آیا پرسپولیس از این مرحله لیگ قهرمانان صعود می‌کند؟
 در مدل آموزش ما همه سندها به بازی دو تیم پرسپولیس و الهلال اشاره دارند

```

# Code
Scores = []
for i in range(len(docs_embedding)):
    Scores.append((i, similarity(docs_embedding[i], query_vec)))
print(len(Scores))
time_finish = datetime.datetime.now()
print("{} results in {} ms".format(len(Scores), ((time_finish - time_start).total_seconds()*1000)))
Scores = sorted(Scores, key = lambda x: x[1], reverse=True)
Show_res_point(Scores , K=5)

# Text
enter a word for checking: آیا پرسپولیس از این مرحله لیگ قهرمانان صعود می‌کند؟
Normal word to search: آیا پرسپولیس از این مرحله لیگ قهرمانان صعود می‌کند؟
{'پرسپولیس': 1, 'مرحله': 1, 'لیگ': 1, 'قهرمان': 1, 'صعود': 1}
7562
7562 results in 101.342 ms
id -> title

The Five Top Related Docs

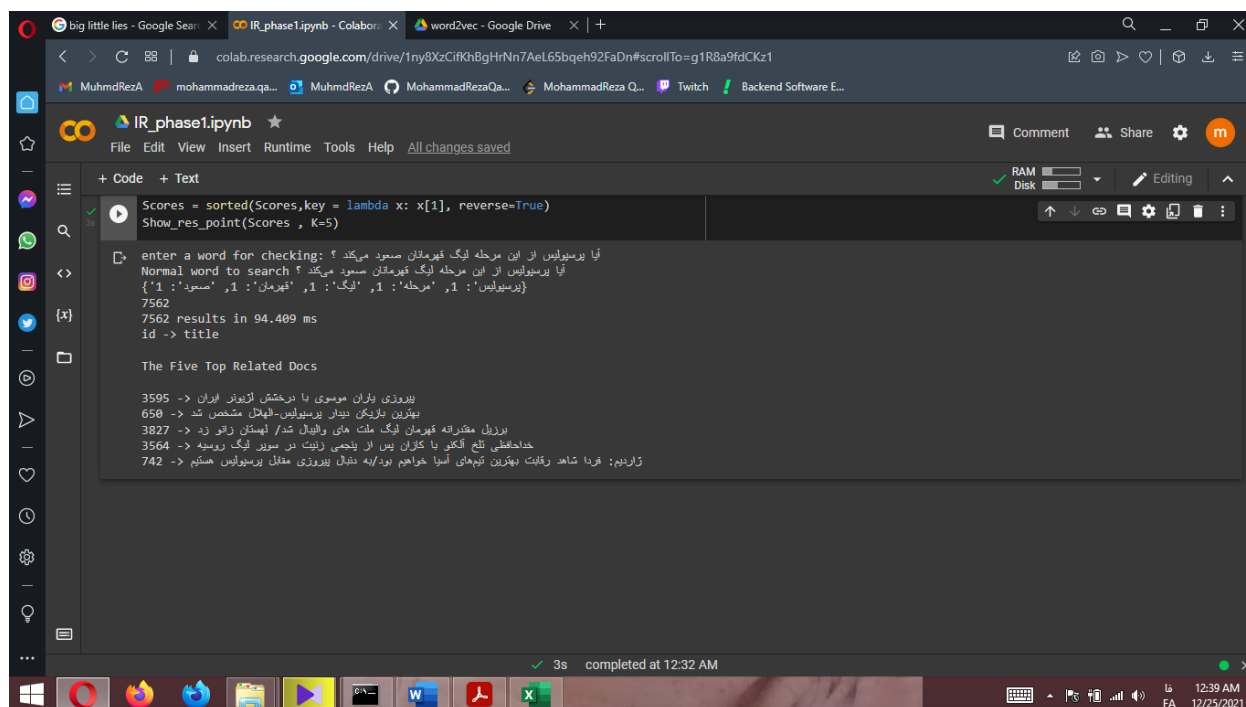
661 -> النصر مدافعتی را برای نیمه نهایی لیگ قهرمانان آسیا از دست داد
742 -> ژاردیم: فردا شاهد رقابت بهترین تیم‌های آسیا خواهیم بود/به دنبال پیروزی مقابل پرسپولیس هستیم
766 -> الهلال با امتیاز متفاوت امتیاز 71 درستی مقابل پرسپولیس+عکس
649 -> به برد قاطع الهلال مقابل پرسپولیس+عکس AFC راکتس جالب
669 -> حضور هواداران الهلال در ورزشگاه العار مقابل پرسپولیس+فیلم
    
```

(اولین داک باز گردانده شده مرتبط است) $Precision@K = 1$

(هر ۵ داک مرتبط اند پس ۱ می‌شود) $Precision@K = 1$

هم ۱ می‌شود $MAP(0.2(5))$

ولی در مدل آموزشی کلی سه تا از اسناد آمده مربوط به لیگ قهرمانان والیبال مربوط می‌شود



```
Scores = sorted(Scores, key = lambda x: x[1], reverse=True)
Show_res_point(Scores , K=5)
```

enter a word for checking: ؟ این مرحله لیگ قهرمانان صعود می‌کند ؟
فیا پرسپولیس از این مرحله لیگ قهرمانان صعود می‌کند ؟
Normal word to search ؟
{ 'پرسپولیس' : 1 , 'مرحله' : 1 , 'لیگ' : 1 , 'قهرمان' : 1 , 'صعود' : 1 }
7562
7562 results in 94.409 ms
id -> title

The Five Top Related Docs

3595 -> پیروزی پارس موسوی با درخشش لژیونر ایران
650 -> بهترین بازیکن دیدار پرسپولیس-الهلال مشخص شد
3827 -> برزیل مقترانه قهرمان لیگ ملت های والیبال شد/ لهستان زانو زد
3564 -> خداحافظی تلخ الکتو با کازان پس از پنجمی زینت در سیزر لیگ روسیه
742 -> زاردم: فردا شاهد رقابت بهترین تیم‌های آسیا خواهیم بود/به دنبال پیروزی مقلد پرسپولیس هستیم

به دلیل اینکه مقادیر بیشتری رو دیده حالت کلی تری در نظر می‌گیرد

(اولین داک باز گردانده شده کاملاً مرتبط نیست) $Precision@K = 0$

(سه تا از ۵ داک غیر مرتبط اند پس 0.4 می‌شود) $Precision@K = 0.4$

هم 0.45 می‌شود $MAP (0.45(0.4 + 0.5))$

حالا برای MRR

برای مدل که خودمون آموزش دادیم

سه تا از کوئری ها جواب درست را در اولین داک برگرداندن از چهارتا پس mrr که میانگین اینکه در چه رنکی بازگشته سند مرتبط هست می‌شود 0.75

برای مدل آموزش داده شده

یکی از کوئری ها پاسخ درست در مکان دوم و دوتا در مکان اول هست و میانگین در این حالت 0.625 می‌شود ولی باید این مورد را در نظر گرفت که این تعداد برای مقایسه این دو کافی نیست

برای برچسب گذاری می‌توانیم از میزان شباهت استفاده کنیم و ترشهولدی در نظر بگیریم.