



دانشگاه صنعتی امیرکبیر  
( پلی تکنیک تهران )

دانشکده مهندسی کامپیوتر

درس بازیابی اطلاعات

## گزارشکار پروژه

موظان میرجلیلی

فرحان کیهان

استاد

دکتر نیک آبادی

زمستان ۱۴۰۱

## فاز ۱

۱) با ذکر مثال شرح دهید که در گام پیش‌پردازش چه عملیاتی انجام داده‌اید. همچنین دلیل انجام هر پردازش را ذکر کنید.

در پیش‌پردازش ابتدا دیتافریم را خواندیم و transpose کردیم تا به فرمت مناسب (هر ستون یک فیچر و هر سطر یک داده) در بیاید. سپس normalize می‌کنیم، در این بخش نیم فاصله‌ها اصلاح می‌شوند و همه کلمات به یک حالت درمی‌آیند. برای مثال "تحریم ها"، "تحریم‌ها" می‌شود. در مرحله بعد یک سری موارد از متن حذف می‌شوند (''< .?@#\$%^&\*~'' و '\n', ' ', '\t', '\r', '\f', '\e')

برای مثال هر کدام از این موارد باشند بعد از این مرحله حذف می‌شوند. tokenization

است که کلمات یک متن به صورت کلمه کلمه و یک لیست در می‌آیند برای مثال عبارت تحریم آمریکا می‌شود یک لیست شامل کلمات تحریم و آمریکا. سپس با استفاده از کتابخانه stopwords از hazm، کلماتی نظیر و، در، از، را و غیره را حذف می‌کنیم که ما در اینجا برای عملکرد بهتر، های و ها هم به این لیست اضافه کرده و از توکنها حذف می‌کنیم. مرحله آخر هم stemming است که کلمات هم‌ریشه و مشابه به یک حالت درمی‌آیند.

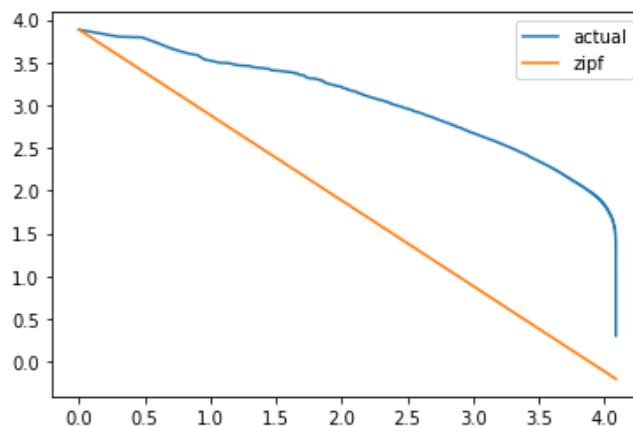
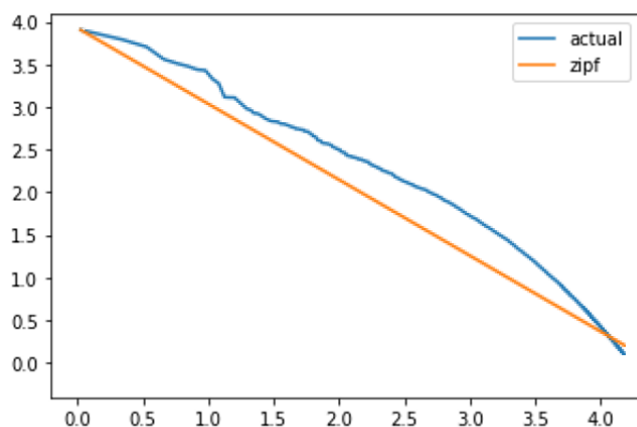
برای مثال خبرگزاری، خبر گزار می‌شود. همه موارد ذکر شده جز punctuation list با استفاده از hazm اجرا شدند.

۲) صحت قانون Zipf را در دو حالت قبل و بعد از حذف کلمات پرتکرار از واژه‌نامه بررسی کنید. رسم نمودار برای هر حالت الزامی است. در صورت برقراری / عدم برقراری این قانون در هر حالت، علت را شرح دهید.

Introduction to Information Retrieval Sec. 5.1	Introduction to Information Retrieval Sec. 5.1
<h3>Zipf's law</h3> <ul style="list-style-type: none"> <li>Heaps' law gives the vocabulary size in collections.</li> <li>We also study the relative frequencies of terms.</li> <li>In natural language, there are a few very frequent terms and very many very rare terms.</li> <li><b>Zipf's law:</b> The <math>i^{\text{th}}</math> most frequent term has frequency proportional to <math>1/i</math>.</li> <li><math>cf_i \propto 1/i = K/i</math> where <math>K</math> is a normalizing constant</li> <li><math>cf_i</math> is <b>collection frequency</b>: the number of occurrences of the term <math>t_i</math> in the collection.</li> </ul>	<h3>Zipf consequences</h3> <ul style="list-style-type: none"> <li>If the most frequent term (<i>the</i>) occurs <math>cf_1</math> times <ul style="list-style-type: none"> <li>then the second most frequent term (<i>of</i>) occurs <math>cf_1/2</math> times</li> <li>the third most frequent term (<i>and</i>) occurs <math>cf_1/3</math> times ...</li> </ul> </li> <li>Equivalent: <math>cf_i = K/i</math> where <math>K</math> is a normalizing factor, so <ul style="list-style-type: none"> <li><math>\log cf_i = \log K - \log i</math></li> <li>Linear relationship between <math>\log cf_i</math> and <math>\log i</math></li> </ul> </li> <li>Another power law relationship</li> </ul>

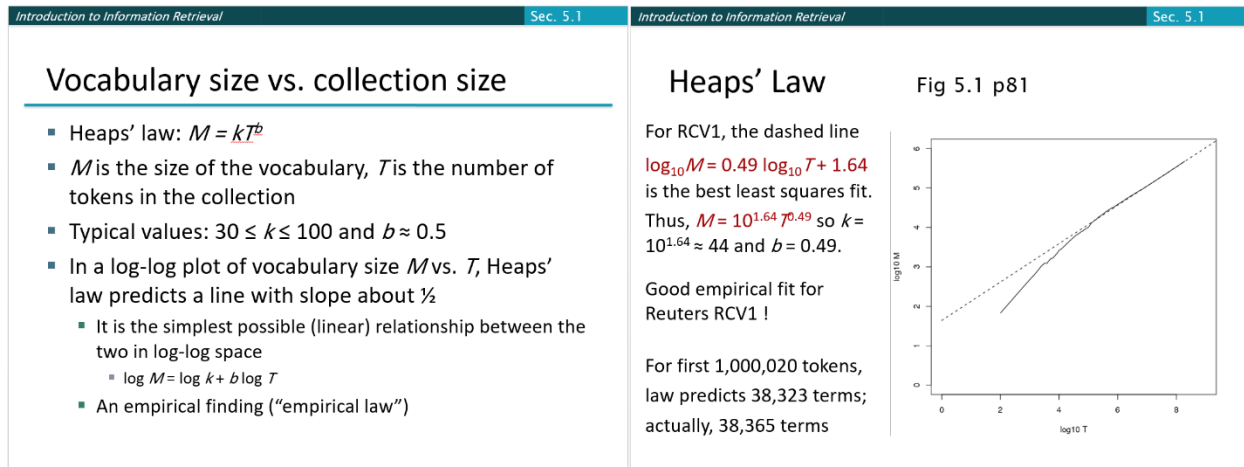
بعد:

قبل:



مشاهده می‌شود که پس از حذف به zipf بیشتر منطبق می‌شود. با توجه به اینکه در بین کلمات ما کلمات پرتکراری وجود دارند که هیچ تاثیری در روند بازیابی ما ندارند با حذف این کلمات به کمک `stopwords_list()` می‌توانیم به شبیه‌تر شدن نمودار ما به مقدار نمودار zipf کمک کند، کلمات ربط، اتصال و ... از کلماتی هستند که در فرمول  $cf/i$  صدق نمی‌کنند و سبب فاصله گرفتن نمودار ما از نمودار zipf می‌شود.

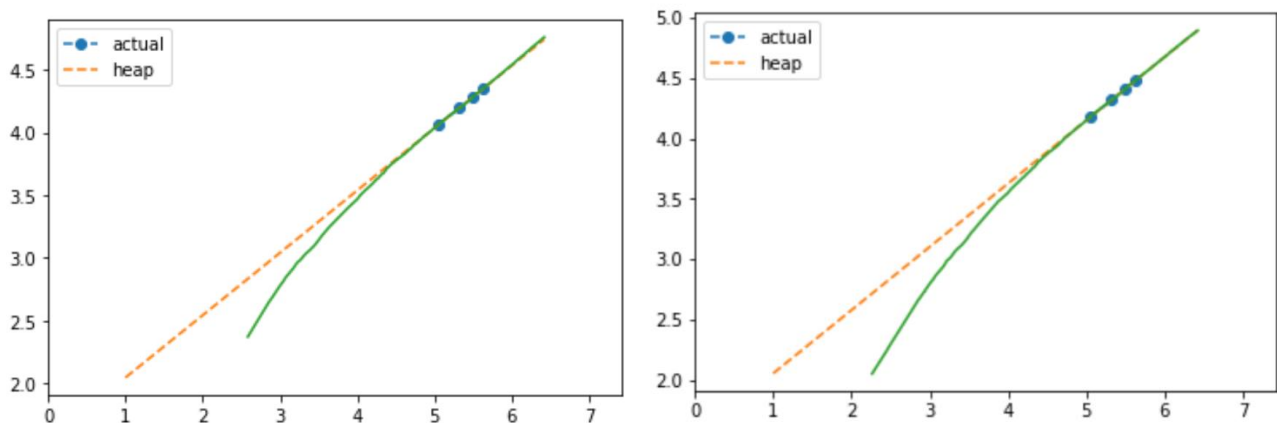
۳) صحت قانون heaps را در دو حالت قبل و بعد از ریشه یابی بررسی کنید. برای بررسی این قانون الزام است با استفاده از اندازه‌ی واژه‌نامه و تعداد توکن‌ها در ۵۰۰، ۱۰۰۰، ۱۵۰۰ و ۲۰۰۰ سند اول، اندازه‌ی واژه‌نامه مربوط به کل اسناد تخمین زده شود. در نهایت اندازه‌ی واقعی واژه‌نامه و اندازه‌ی تخمینی در هر دو حالت مقایسه و تحلیل شود. آیا در هر دو حالت قانون برقرار است؟ چرا؟ رسم نمودار برای هر حالت الزامی است.



در ابتدا می‌بینیم که دیتافریم به دلیل نرمال نبودن توزیع داده‌ها در heaps صدق نمی‌کند اما بعد از ریشه‌یابی می‌بینیم که در قانون heaps صدق می‌کند.

قبل:

بعد:



در ابتدا با قرار دادن پشت هم اسناد ورودی تابع ما دارای انحراف از heap's می‌شود به همین دلیل ما با کمک گرفتن از انتخاب تصادفی اسناد را انتخاب کردیم که در تصاویر بالا مشاهده می‌کنیم. در ابتدا قبا از ریشه‌یابی مشاهده می‌کنیم که تابع ما به خوبی بر روی heap's fit شده است سپس پس از ریشه‌یابی مشاهده می‌کنیم که

باز هم بر روی آن fit می‌شود دلیل این مورد این است که در اینجا ریشه‌یابی کلماتی که از ریشه مشترک هستند را حذف می‌کند ولی در مقابل در فرمول heap's با تغییر ضریب b دوباره بر روی heap's قرار می‌گیرد.

۴) حداقل سه مورد از مواردی که در ریشه‌یابی با چالش روبرو بودید را ذکر کنید. بطور مثال کلماتی که نیازی به ریشه‌یابی ندارند اما طبق روند ریشه‌یابی از دست می‌روند.

در فرایند ریشه‌یابی کلمات فارسی با چالش‌های مختلفی روبه‌رو هستیم که باید با شناسایی آنها به خوبی برای جلوگیری از آنها اقدام کرد. در ادامه تعدادی از چالش‌هایی که وجود دارد را بیان می‌کنیم:

۱. مشکلات موجود در تعیین مرز کلمات : در زبان فارسی، برای افزودن پیشوند و پسوند، در مواردی از فاصله و در مواردی از نیم‌فاصله استفاده می‌شود. بنابراین وجود فاصله به معنای مرز کلمات نیست .
۲. انواع جمع‌ها : جمع‌های مکسر از زبان عربی در زبان فارسی وارد شده‌اند که از قوانین زبان فارسی برای ریشه‌یابی استفاده نمی‌کنند. در واقع این کلمات قاعده خاصی برای جمع زدن ندارند
۳. عدم استفاده از اعراب : عدم استفاده از اعراب گاهی باعث می‌شود که ریشه کلماتی که دارای ریشه‌های متفاوتی هستند، یکسان در نظر گرفته شود. همچنین از آنجایی که یک کلمه بدون اعراب می‌تواند معانی مختلفی داشته باشد، تشخیص فعل از اسم را مشکل می‌نماید.
۵. بی قاعده بودن برخی از افعال و ابهام در تولید ریشه آنها مانند رفتن و رو که در اینجا این مورد در کتابخانه hazm در نظر گرفته شده است.
۶. وجود استثناهای فراوان : بعضی از کلمات فارسی، دارای استثناهایی می‌باشند . همانند <ان>، این < که در بعضی مواقع به عنوان علامت جمع به کار می‌رود ولی در بعضی کلمات متعلق به خود کلمه است. درختان، فراوان-مسافرتین، قزوین

۵) پاسخ به پرسمان در حالت های زیر:

الف) یک پرسمان از کلمات ساده و متداول مانند تحریم‌های آمریکا علیه ایران، در نتایج بازیابی شده انتظار می‌رود اسنادی که کلمات تحریم، آمریکا، علیه و ایران را دارند در بالای لیست و اسنادی که برخی از کلمات را ندارند در رتبه‌های پایینتر لیست قرار داشته باشند.

Output for query: the multi\_word\_query (فدراسیون فوتبال ایران)

[ ] multi\_word\_query(query, must\_query)

```
doc number: 5632 -> title: هفته سیزدهم لیگ برتر فوتبال | تداوم شکست تاجیک‌های استقلال در اراک؟ / نبرد قزمیوشان تهران و تبریز
doc number: 3593 -> title: تیم ملی-لیل-غیت-هشتمین-جست url: https://www.farsnews.ir/news/14001106000795/
doc number: 1548 -> title: زمان معرفی رسمی میزبان لیگ قهرمانان مشخص شد/خوشحالی سوندی‌ها از ناکامی استقلال برای بازگشت به آسیا
doc number: 6165 -> title: فراسون و سازمان لیگ درباره مجوز حرفه‌ای باید خیلی جدی‌تر رفتار می‌کردند
doc number: 6170 -> title: فریدشوران: خوبان عالم کمک کند تا آلفان این همه سوتی ندهند/ فراسون فوتبال شایه کرد
doc number: 1572 -> title: قلم‌نویس: دیدار با پرسپولیس فیلی برای بازگشت ما به کورس است/به جای 7 امتیاز، 15 امتیاز ازما گرفتند
doc number: 3620 -> title: سلطانی: پرسپولیس‌ها تحمل کند تا کاسه درویش پر شود/ بازیکن با اعصاب به جلی نمی‌رسد
doc number: 5160 -> title: محبی: یاد بازی پرسپولیس اقدام/قلعه رخ داده، اما مسیبتش مسئولیت بالاتر گرفته‌اند/فراسون‌مان زور ندارد
doc number: 6697 -> title: جبهانبخش: طارمی گفت برگردان من به چلی بهتر بود/ کپروش روی ذهیت ما خیلی تاثیر گذاشت/ نتایج اسکوچیک کیفیت او را نشان می‌دهد
doc number: 6698 -> title: جایزه مرد سال آسیا از دست آزمون و طارمی برد/خبری از ایرانی‌ها در میان بهترین‌ها نیست+عکس
doc number: 2098 -> title: مصاحبه فارس با کارشناس فوتبال آسیا | از میزبان بزرگ کپروش و قدرت ایران با اسکوچیک تا انقلاب برانکو در صن
doc number: 3141 -> title: کپنه استیلف رتی بازیکن گابنی گل گهر را تأیید کرد/ امتیازها به تیم سیرجانی برگردانده نمی‌شود
doc number: 80 -> title: اقدام فراسون فوتبال برای اجرای شن کردن اعضای کادر فنی تیم ملی و پرداخت مطالبات
doc number: 84 -> title: ناظم-الشریعه url: https://www.farsnews.ir/news/14001223000360/ تمجید ویژه رئیس فدراسیون عراق از ناظم الشریعه
doc number: 3156 -> title: حمید اسبلی-سفر-ورزش-هنگای-شد url: https://www.farsnews.ir/news/14001112000513/ حمید اسبلی سفر ورزش هنگای شد
```

(ب) یک پرسمان با عملگر NOT مانند تحریم‌های آمریکا! ایران، انتظار می‌رود اسنادی که شامل دو کلمه تحریم و آمریکا هستند اما کلمه‌ی ایران را ندارند در نتایج باز یابی شده وجود داشته باشند.

Output for query: {تحریم‌های آمریکا! ایران} using the multi\_word\_query

In this scenario we must handle the "!" as not for the word after this sign and then aggregate docs

- ▶ `multi_word_query(query, must_query)`

doc number: 9984 -> title: <https://www.farsnews.ir/news/14000812000209> url: <https://www.farsnews.ir/news/14000812000209> doc number: 11393 -> title: <https://www.farsnews.ir/news/14000801000800> url: <https://www.farsnews.ir/news/14000801000800> doc number: 11918 -> title: <https://www.farsnews.ir/news/14001097000282> url: <https://www.farsnews.ir/news/14001097000282> doc number: 9361 -> title: <https://www.farsnews.ir/news/1400128000581> url: <https://www.farsnews.ir/news/1400128000581> doc number: 7315 -> title: <https://www.farsnews.ir/news/14001222000366> url: <https://www.farsnews.ir/news/14001222000366> doc number: 6942 -> title: <https://www.farsnews.ir/news/140010970008215> url: <https://www.farsnews.ir/news/140010970008215> doc number: 9377 -> title: <https://www.farsnews.ir/news/14001212000557> url: <https://www.farsnews.ir/news/14001212000557> doc number: 7084 -> title: <https://www.farsnews.ir/news/14001212000557> url: <https://www.farsnews.ir/news/14001212000557> doc number: 6959 -> title: <https://www.farsnews.ir/news/14001221001176> url: <https://www.farsnews.ir/news/14001221001176> doc number: 7346 -> title: <https://www.farsnews.ir/news/14001289000476> url: <https://www.farsnews.ir/news/14001289000476> doc number: 7230 -> title: <https://www.farsnews.ir/news/1400124000432> url: <https://www.farsnews.ir/news/1400124000432> doc number: 7615 -> title: <https://www.farsnews.ir/news/14001129000373> url: <https://www.farsnews.ir/news/14001129000373> doc number: 9280 -> title: <https://www.farsnews.ir/news/14001008000738> url: <https://www.farsnews.ir/news/14001008000738> doc number: 11599 -> title: <https://www.farsnews.ir/news/14000811000708> url: <https://www.farsnews.ir/news/14000811000708> doc number: 7261 -> title: <https://www.farsnews.ir/news/14001212000783> url: <https://www.farsnews.ir/news/14001212000783>

پ) یک پرسمان با عملگر عبارت مانند "کنگره ضدتروریست"، انتظار می‌رود اسنادی که شامل عبارت کنگره ضدتروریست در نتایج بازبایی شده وجود داشته باشند؛ به عبارت دیگر موقعیت مکانی کلمات در این حالت مهم است.

▶ `multi_word_query(query, must_query)`

executed by Farhan kayhan  
8:41 PM (1 minute ago)  
executed in 0.007s

ت) یک پرسمان پیچیده مانند " تحریم هسته‌ای " آمریکا ! ایران، انتظار می‌رود اسنادی که شامل عبارت تحریم هسته‌ای و کلمه‌ی آمریکا هستند اما کلمه‌ی ایران را ندارند در نتایج بازایی شده وجود داشته باشد.

Output for query: "تحریم هسته ای"آمریکا!ایران using the multi\_word\_query

[ ] multi\_word\_query(query, must\_query)

doc number: 11393 -> title: امروز-طرف اصلی/ <https://www.farsnews.ir/news/14000812000209> url: <https://www.farsnews.ir/news/14000812000209>  
doc number: 7395 -> title: دانشگاه تهران: رویکرد «ماتر» تجربه تلخ برجام را نصیب تاریخ پرافتخار انقلاب کرد <https://www.farsnews.ir/news/1400120700>  
doc number: 10862 -> title: دانشگاهی به بهانه آبان، برای پرونده‌سازی حقوق بشری/داده‌هایی که هیچ‌گاه تشکیل نمی‌شود <https://www.farsnews.ir/news/14000826000161>  
doc number: 7346 -> title: تجمع دانشجویان در فروگاه امام| تاکید بر رعایت شروط رهبر انقلاب در مذاکرات <https://www.farsnews.ir/news/14001209000476/>  
doc number: 7315 -> title: جامعه اسلامی دانشجویان: غربی‌ها ثابت کرده‌اند تصمیمی برای لغو تحریم‌ها ندارند <https://www.farsnews.ir/news/14001210000581>  
doc number: 10162 -> title: روایت خبرنگار فارس از یک همایش دانشجویی/ 2 ساعت و نیم صریح و بدون تعارف با رئیس مجلس <https://www.farsnews.ir/news/14000915000494/>  
doc number: 9980 -> title: ضلع تفکیک-نایبیر- از-مذاکرات <https://www.farsnews.ir/news/14000920000012/> url: <https://www.farsnews.ir/news/14000920000012/>  
doc number: 7615 -> title: چنان‌زنی-مسئولان-در-مذاکرات-بود <https://www.farsnews.ir/news/14001129000373/> url: <https://www.farsnews.ir/news/14001129000373/>

ث) یک پرسمان کلمات نادر مانند اورشلیم ! صهیونیست، خروجی مورد انتظار این قسمت مشابه با قسمت ب می‌باشد با این تفاوت که کلمات استفاده شده در پرسمان از کلمات نادر هستند.

[136] multi\_word\_query(query, must\_query)

'no related query found'

مشاهده می‌شود در این صورت هیچ سندی برای کلمه اورشلیم به تنهایی وجود ندارد.

## فاز ۲

در این بخش قصد داریم تا با به دست آوردن امتیاز و محاسبه معیار شباهت اسناد برای یک کوئری ورودی بهترین اسناد بر برای کاربر استخراج کنیم.

در ابتدا با محاسبه tf.idf برای اسناد و سپس محاسبه tf برای کوئری به معیار امتیازدهی مدنظر خود می‌رسیم، سپس با نرمالایز کردن امتیاز به دست آمده بر نرم ۲ بردار اسناد و کوئری به کسینوس بین دو بردار می‌رسیم. به این معیار به دست آمده شباهت کسینوسی گفته می‌شود.

در بخش بعد با در نظر گرفتن معیارهای به دست آمده یک champion list برای بهترین اسناد هر کدام از کلمه‌های خود درست می‌کنیم، این کار کمک می‌کند تا بتوانیم در مواقع مواجهه با یک کوئری از بین بهترین اسناد مرتبط با آن کوئری اسناد را استخراج کنیم و همچنین به دلیل اینکه اسناد از پیش آماده هستند با سرعت پردازش سریع‌تری مواجه هستیم ولی از طرفی این یک روش lossy هست و لزوماً top-k اصلی را به ما نمی‌دهد ولی سرعت پاسخ‌گویی را چند برابر می‌کند.



## (۱) پاسخ به پرسمان در حالت های زیر:

الف) با توجه به کلمه ساده انتخاب شده که «فدراسیون» است مشاهده می کنیم که در بخش اول اسنادی که وزن بیشتری از آن کلمه را دارند به ما نمایش داده می شود ولی در champion list در واقع نتایج بهتری داریم زیرا کلمه فدراسیون می تواند استفاده هایی خاص منظوره ای در اسناد ورزشی داشته باشد و همانطور که مشاهده می کنیم روش دوم نتایج مطلوب تری را برای ما کسب می کند.

```
One word common query -> "فدراسیون"

cosine_similarity(query)

<ipython-input-49-3fcf7064d513>:10: RuntimeWarning: invalid value encountered in double_scalars
score=temp/(np.linalg.norm(list(query_tf(query).values()))*norm2_doc(query, i))
doc number: 12118 -> title: url: https://www.farsnews.ir/news/140007260005
doc number: 2482 -> title: url: https://www.farsnews.ir/news/14001120000902/
doc number: 2437 -> title: url: https://www.farsnews.ir/news/14001121000653/
doc number: 2449 -> title: url: https://www.farsnews.ir/news/14001121000484/
doc number: 2453 -> title: url: https://www.farsnews.ir/news/14001119001101/
doc number: 2458 -> title: url: https://www.farsnews.ir/news/14001121000288/
doc number: 2460 -> title: url: https://www.farsnews.ir/news/140011210001064/
doc number: 2464 -> title: url: https://www.farsnews.ir/news/14001121000206/
doc number: 2475 -> title: url: https://www.farsnews.ir/news/14001121000077/
doc number: 2478 -> title: url: https://www.farsnews.ir/news/14001120001064/
doc number: 2483 -> title: url: https://www.farsnews.ir/news/14001120000968/
doc number: 2429 -> title: url: https://www.farsnews.ir/news/14001121000077/
doc number: 2486 -> title: url: https://www.farsnews.ir/news/14001120000906/
doc number: 2488 -> title: url: https://www.farsnews.ir/news/14001120000886/
doc number: 2491 -> title: url: https://www.farsnews.ir/news/14001120000841/
doc number: 2495 -> title: url: https://www.farsnews.ir/news/14001120000780/
doc number: 2500 -> title: url: https://www.farsnews.ir/news/14001120000720/
doc number: 2501 -> title: url: https://www.farsnews.ir/news/14001120000710/
doc number: 2503 -> title: url: https://www.farsnews.ir/news/14001120000684/
```

```
champion_similarity(query)

<ipython-input-55-97d8424d142a>:10: RuntimeWarning: invalid value encountered in double_scalars
score=temp/(np.linalg.norm(list(query_tf(query).values()))*norm2_doc(query, i))
doc number: 6128 -> title: url: https://www.farsnews.ir/news/14001021000577/
doc number: 4918 -> title: url: https://www.farsnews.ir/news/14001121000699/
doc number: 3057 -> title: url: https://www.farsnews.ir/news/14001121000699/
doc number: 562 -> title: url: https://www.farsnews.ir/news/14001121000699/
doc number: 2336 -> title: url: https://www.farsnews.ir/news/14001121000699/
doc number: 4873 -> title: url: https://www.farsnews.ir/news/14001121000699/
doc number: 3073 -> title: url: https://www.farsnews.ir/news/14001121000699/
doc number: 544 -> title: url: https://www.farsnews.ir/news/14001121000699/
doc number: 1403 -> title: url: https://www.farsnews.ir/news/14001121000699/
doc number: 3082 -> title: url: https://www.farsnews.ir/news/14001121000699/
doc number: 2268 -> title: url: https://www.farsnews.ir/news/14001121000699/
doc number: 2264 -> title: url: https://www.farsnews.ir/news/14001121000699/
doc number: 4907 -> title: url: https://www.farsnews.ir/news/14001121000699/
doc number: 3141 -> title: url: https://www.farsnews.ir/news/14001121000699/
doc number: 2253 -> title: url: https://www.farsnews.ir/news/14001121000699/
doc number: 469 -> title: url: https://www.farsnews.ir/news/14001121000699/
doc number: 2245 -> title: url: https://www.farsnews.ir/news/14001121000699/
doc number: 3045 -> title: url: https://www.farsnews.ir/news/14001121000699/
doc number: 458 -> title: url: https://www.farsnews.ir/news/14001121000699/
doc number: 5590 -> title: url: https://www.farsnews.ir/news/14001121000699/
```

ب) در بخش بعدی کلمه انتخاب شده «فدراسیون فوتبال» است و همانطور که مشاهده می‌کنیم در اینجا هر دو روش نتایج قابل قبول مرتبط با کوئری وارد شده کسب می‌کنند ولی در زمان اجرا champion list با سرعت بالاتری اطلاعات را برای ما نمایش می‌دهد زیرا اطلاعات مرتبط به آن از پیش آماده هستند.

```
Multiple word common query -> "فدراسیون فوتبال"

[ ] cosine_similarity(query)

<ipython-input-49-3fcf7064d513>:10: RuntimeWarning: invalid value encountered in double_scalars
score=temp/(np.linalg.norm(list(query_tf(query).values()))*norm2_doc(query, i))
doc number: 6327 -> title: url: https://www.farsnews.ir/news/14001002000210/ لیس-به-آسیا/
doc number: 1463 -> title: url: https://www.farsnews.ir/news/14001204001205/ تده-با-تمام/
doc number: 5285 -> title: url: https://www.farsnews.ir/news/14001017000075/ باشدل-حتما/
doc number: 555 -> title: url: https://www.farsnews.ir/news/14001216000770/
doc number: 5972 -> title: url: https://www.farsnews.ir/news/14001007000268/ ملی/
doc number: 3510 -> title: url: https://www.farsnews.ir/news/14001111000111/
doc number: 3479 -> title: url: https://www.farsnews.ir/news/14001107000765/
doc number: 3064 -> title: url: https://www.farsnews.ir/news/14001113000359/
doc number: 2488 -> title: url: https://www.farsnews.ir/news/14001120000886/
doc number: 6109 -> title: url: https://www.farsnews.ir/news/14001005000211/
doc number: 1585 -> title: url: https://www.farsnews.ir/news/14001203000743/
doc number: 3194 -> title: url: https://www.farsnews.ir/news/14001111000488/
doc number: 5403 -> title: url: https://www.farsnews.ir/news/14001014001407/
doc number: 8828 -> title: url: https://www.farsnews.ir/news/14001021000830/
doc number: 4617 -> title: url: https://www.farsnews.ir/news/14001025000252/
doc number: 6134 -> title: url: https://www.farsnews.ir/news/14001004000651/
doc number: 1553 -> title: url: https://www.farsnews.ir/news/14001204000081/
doc number: 409 -> title: url: https://www.farsnews.ir/news/14001218000584/
doc number: 4315 -> title: url: https://www.farsnews.ir/news/14001028000236/

[ ] champion_similarity(query)

<ipython-input-55-97d8424d142a>:10: RuntimeWarning: invalid value encountered in double_scalars
score=temp/(np.linalg.norm(list(query_tf(query).values()))*norm2_doc(query, i))
doc number: 6128 -> title: url: https://www.farsnews.ir/news/14001021000577/
doc number: 4918 -> title: url: https://www.farsnews.ir/news/14001021000577/
doc number: 1950 -> title: url: https://www.farsnews.ir/news/14001128000554/
doc number: 2268 -> title: url: https://www.farsnews.ir/news/14001128000554/
doc number: 6323 -> title: url: https://www.farsnews.ir/news/14001128000554/
doc number: 5102 -> title: url: https://www.farsnews.ir/news/14001019000346/
doc number: 139 -> title: url: https://www.farsnews.ir/news/14001222000297/
doc number: 5583 -> title: url: https://www.farsnews.ir/news/14001012000666/
doc number: 4603 -> title: url: https://www.farsnews.ir/news/14001025000363/
doc number: 3057 -> title: url: https://www.farsnews.ir/news/140011130006/
doc number: 1824 -> title: url: https://www.farsnews.ir/news/14001130000599/
doc number: 3457 -> title: url: https://www.farsnews.ir/news/14001130000599/
doc number: 3141 -> title: url: https://www.farsnews.ir/news/14001112000714/
doc number: 6764 -> title: url: https://www.farsnews.ir/news/14001112000714/
doc number: 5252 -> title: url: https://www.farsnews.ir/news/14001112000714/
doc number: 2626 -> title: url: https://www.farsnews.ir/news/14001112000714/
doc number: 7305 -> title: url: https://www.farsnews.ir/news/14001210000621/
doc number: 5221 -> title: url: https://www.farsnews.ir/news/14001018000261/
doc number: 5005 -> title: url: https://www.farsnews.ir/news/14001020000261/
doc number: 1974 -> title: url: https://www.farsnews.ir/news/14001020000261/
```

پ) کلمه سخت انتخاب شده ما «گابن» است. در سال گذشته بازی کردن یک بازیکن گابنی غیرمجاز در لیگ فوتبال ایران تا مدت‌ها سوژه خبرگزاری‌های داخلی بود به همین دلیل انتظار می‌رود که با وارد کردن این عبارت به اسناد مرتبط از دسته ورزشی برخورد کنیم که مشابه آنچه در تصویر می‌بینید تمامی اسناد استخراج شده مرتبط با همین موضوع هستند.

```

One rare word -> "گاین"
[ ] cosine_similarity(query)

<ipython-input-49-3fcf7064d513>:10: RuntimeWarning: invalid value encountered in double_scalars
score=temp/(np.linalg.norm(list(query_tf.query).values()))*norm2_doc(query,i))
doc number: 6799 -> title: https://www.farsnews.ir/news/140011110000661 url: https://www.farsnews.ir/news/140011110000661
doc number: 3717 -> title: https://www.farsnews.ir/news/140011110000663 url: https://www.farsnews.ir/news/140011110000663
doc number: 3242 -> title: https://www.farsnews.ir/news/140011110000540 url: https://www.farsnews.ir/news/140011110000540
doc number: 3236 -> title: https://www.farsnews.ir/news/140011110000332 url: https://www.farsnews.ir/news/140011110000332
doc number: 3455 -> title: https://www.farsnews.ir/news/140011027000456 url: https://www.farsnews.ir/news/140011027000456
doc number: 3457 -> title: https://www.farsnews.ir/news/140011027000456 url: https://www.farsnews.ir/news/140011027000456
doc number: 3807 -> title: https://www.farsnews.ir/news/140011027000456 url: https://www.farsnews.ir/news/140011027000456
doc number: 3141 -> title: https://www.farsnews.ir/news/140011030000767 url: https://www.farsnews.ir/news/140011030000767
doc number: 3894 -> title: https://www.farsnews.ir/news/140011030000767 url: https://www.farsnews.ir/news/140011030000767
doc number: 4048 -> title: https://www.farsnews.ir/news/140011030000767 url: https://www.farsnews.ir/news/140011030000767
doc number: 4138 -> title: https://www.farsnews.ir/news/140011030000767 url: https://www.farsnews.ir/news/140011030000767
doc number: 4419 -> title: https://www.farsnews.ir/news/140011027000456 url: https://www.farsnews.ir/news/140011027000456
doc number: 4459 -> title: https://www.farsnews.ir/news/140011026000849 url: https://www.farsnews.ir/news/140011026000849
doc number: 4564 -> title: https://www.farsnews.ir/news/140011027000456 url: https://www.farsnews.ir/news/140011027000456
doc number: 3203 -> title: https://www.farsnews.ir/news/140011110000887 url: https://www.farsnews.ir/news/140011110000887
doc number: 3103 -> title: https://www.farsnews.ir/news/14001112001965 url: https://www.farsnews.ir/news/14001112001965
doc number: 6764 -> title: https://www.farsnews.ir/news/140011027000456 url: https://www.farsnews.ir/news/140011027000456
doc number: 1709 -> title: https://www.farsnews.ir/news/140011027000456 url: https://www.farsnews.ir/news/140011027000456

```

```
[ ] champion_similarity(query)

cipython-input-55-97d8424d142a>:10 RuntimeWarning: invalid value encountered in double_scalars
score=tmf(np.linalg.norm(linalg.norm(tmf(query).values()))*norm2(doc(query), i))
doc number: 6799 -> title: https://www.farsnews.ir/news/14001110006603: اعتراف سوریمی پرنکیله بریوئیل و سورینگ لیس سزاد بچنی/تکرار حرکت مجیدی توسط قهوه نبوی 8
doc number: 3717 -> title: https://www.farsnews.ir/news/14001110006603: از مشکلات شاعرانه تا کلیه ستاره میباریزی پرسپولیس / در انتظار شوک محرم، کینه مسابقات مرودند 8
doc number: 3224 -> title: https://www.farsnews.ir/news/14001110006603: رای-پیونده بازیکن گانی، گرگیور مشخص شد/ url:
doc number: 3232 -> title: https://www.farsnews.ir/news/14001110006603: خرید-جید-ژای-به شهر-پارسلونا رسید-عکس/ url: خرید چند ژای به شهر پارسلونا رسید عکس
doc number: 3236 -> title: https://www.farsnews.ir/news/14001110006603: ی-فروش دمیله- ستاره فراسوی ماندن-شد/ url: پارسلونا در تکتی فروش دمیله / ستاره فراسوی منتفی شد؟
doc number: 3257 -> title: https://www.farsnews.ir/news/14001110006603: ۲۵-۱۴: پیونده بازیکن گل گیر به روزهای آخر رسید/دارس-من گزارش 2 تصصیفاتی را پلیس تا خدمت
doc number: 3445 -> title: https://www.farsnews.ir/news/14001108000467: خرید-جید-پارسلونا کیست/ url: خرید چند بازیکن کیست؟
doc number: 3457 -> title: https://www.farsnews.ir/news/14001104000123: ای-گانی- مارا به داوگری متهم کردن ولی چلوگری شرکت است/خونوز رای کیوش ملیط میدهم/مگر اسکوچیچ گرفت؟
doc number: 3807 -> title: https://www.farsnews.ir/news/1400112000715: ای-گانی- مارا به داوگری متهم کردن ولی چلوگری شرکت است/خونوز رای کیوش ملیط میدهم/مگر اسکوچیچ گرفت؟
doc number: 3141 -> title: https://www.farsnews.ir/news/14001103000320: ی-بیت-بیت-فکری به-آرسلونا رسید/ url: کینه استیف رای بازیکن گانی گل گیر را تأیید کرد/ امتیازها به تیم سرچایی برگردانده نمی‌شود
doc number: 3894 -> title: https://www.farsnews.ir/news/14001103000320: ی-بیت-بیت-فکری به-آرسلونا رسید/ url: کینه استیف رای بازیکن گانی گل گیر را تأیید کرد/ امتیازها به تیم سرچایی برگردانده نمی‌شود
doc number: 4040 -> title: https://www.farsnews.ir/news/14001026000849: ی-بیت-بیت-فکری به-آرسلونا رسید/ url: کینه استیف رای بازیکن گانی گل گیر را تأیید کرد/ امتیازها به تیم سرچایی برگردانده نمی‌شود
doc number: 4138 -> title: https://www.farsnews.ir/news/14001026000849: ی-بیت-بیت-فکری به-آرسلونا رسید/ url: کینه استیف رای بازیکن گانی گل گیر را تأیید کرد/ امتیازها به تیم سرچایی برگردانده نمی‌شود
doc number: 4419 -> title: https://www.farsnews.ir/news/14001026000849: ی-بیت-بیت-فکری به-آرسلونا رسید/ url: کینه استیف رای بازیکن گانی گل گیر را تأیید کرد/ امتیازها به تیم سرچایی برگردانده نمی‌شود
doc number: 4459 -> title: https://www.farsnews.ir/news/14001026000849: ی-بیت-بیت-فکری به-آرسلونا رسید/ url: کینه استیف رای بازیکن گانی گل گیر را تأیید کرد/ امتیازها به تیم سرچایی برگردانده نمی‌شود
doc number: 4564 -> title: https://www.farsnews.ir/news/14001026000849: ی-بیت-بیت-فکری به-آرسلونا رسید/ url: کینه استیف رای بازیکن گانی گل گیر را تأیید کرد/ امتیازها به تیم سرچایی برگردانده نمی‌شود
doc number: 3203 -> title: https://www.farsnews.ir/news/14001103000320: ی-بیت-بیت-فکری به-آرسلونا رسید/ url: کینه استیف رای بازیکن گانی گل گیر را تأیید کرد/ امتیازها به تیم سرچایی برگردانده نمی‌شود
doc number: 3103 -> title: https://www.farsnews.ir/news/1400112001065/: حضور-ستاره-در-است-داد/ url: ستاره بزرگ به ستاره اخراجی آرسل / اولیمپیک جام ملت ها را از دست داد؟
doc number: 6764 -> title: https://www.farsnews.ir/news/14001026000849: ی-بیت-بیت-فکری به-آرسلونا رسید/ url: کینه استیف رای بازیکن گانی گل گیر را تأیید کرد/ امتیازها به تیم سرچایی برگردانده نمی‌شود
doc number: 1709 -> title: https://www.farsnews.ir/news/14001201001230: ی-بیت-بیت-فکری به-آرسلونا رسید/ url: کینه استیف رای بازیکن گانی گل گیر را تأیید کرد/ امتیازها به تیم سرچایی برگردانده نمی‌شود
```

ت) در این بخش با ترکیب کلمات «فدراسیون فوتبال گابن» می‌خواهیم اسناد بازگشتی را مشاهده کنیم. ما در بین اسناد، اسنادی را موجود داریم که عیناً ترکیب این ۳ کلمه را دارند ولی در اینجا به دلیل اینکه تنها تکرار کلمات مختلف از اهمیت بالایی برخوردار است نمی‌توانیم در واقع مرتبط‌ترین سند را که شامل این کلمه است را مشاهده کنیم در صورتی که با استفاده از positional index می‌توان سند مرتبط را دریافت کرد.

Multiple rare words —> "فراسيون فوتبال گابن"

▶ cosine\_similarity(query)

```

cypthon-input-49-3fcf70645d310: RuntimeWarning: invalid value encountered in double_scalars
score=temp/(np.linalg.norm(list((query_tf(query).values()))*norm2.doc(query, i)))
doc number: 6327 -> title: url: https://www.farsnews.ir/news/14001002000210 سولیس: به این/
doc number: 1797 -> title: https://www.farsnews.ir/news/14001002000210 رکن قضایی بعد از اعتراض اول به بازبینی گانی بد دستور فوق‌العاده داد/واکنش جالب است به وصفه ورنش
doc number: 3048 -> title: https://www.farsnews.ir/news/14001130005111 اخبار: بنگاه‌ها ورود/14001130005111 حق صون کینه استیضایی درخصوصی فوق‌العاده: به بی‌بهره‌ای اخیر بنگاه‌ها ورنش
doc number: 3057 -> title: https://www.farsnews.ir/news/14001130004361 اخبار: بنگاه‌ها: به تصاحیح کینه: به بی‌بهره‌ای اخیر بنگاه‌ها ورنش
doc number: 4661 -> title: https://www.farsnews.ir/news/14001024000169 ر.ه.ای: استکل/14001024000169 سکت: وزارت ورنش به سرجی با تفصیل موضوعی درباره/ای استکل اسبلی کنده
doc number: 5377 -> title: https://www.farsnews.ir/news/1400101030003201 اخبار: ر.ه.ای: استکل/1400101030003201 حق صون کینه استیضایی درخصوصی فوق‌العاده: به بی‌بهره‌ای اخیر بنگاه‌ها ورنش
doc number: 5373 -> title: https://www.farsnews.ir/news/14001015000545 اخبار: ر.ه.ای: استکل/14001015000545 رنیش کینه استیضایی: زمان تصحیح ورنش در مورد برونه گنی مشخص نیست/اشاری روی ما کرده استند
doc number: 3894 -> title: https://www.farsnews.ir/news/140011030003201 اخبار: ر.ه.ای: استکل/140011030003201 حق صون کینه استیضایی درخصوصی فوق‌العاده: به بی‌بهره‌ای اخیر بنگاه‌ها ورنش
doc number: 3457 -> title: https://www.farsnews.ir/news/1400101030003201 اخبار: ر.ه.ای: استکل/1400101030003201 حق صون کینه استیضایی درخصوصی فوق‌العاده: به بی‌بهره‌ای اخیر بنگاه‌ها ورنش
doc number: 6700 -> title: https://www.farsnews.ir/news/140000927000277 اخبار: ر.ه.ای: استکل/140000927000277 حق صون کینه استیضایی درخصوصی فوق‌العاده: به بی‌بهره‌ای اخیر بنگاه‌ها ورنش
doc number: 6566 -> title: https://www.farsnews.ir/news/1400101030003201 اخبار: ر.ه.ای: استکل/1400101030003201 حق صون کینه استیضایی درخصوصی فوق‌العاده: به بی‌بهره‌ای اخیر بنگاه‌ها ورنش
doc number: 3101 -> title: https://www.farsnews.ir/news/1400101030003201 اخبار: ر.ه.ای: استکل/1400101030003201 حق صون کینه استیضایی درخصوصی فوق‌العاده: به بی‌بهره‌ای اخیر بنگاه‌ها ورنش
doc number: 1572 -> title: https://www.farsnews.ir/news/1400102000104 اخبار: ر.ه.ای: استکل/1400102000104 حق صون کینه استیضایی درخصوصی فوق‌العاده: به بی‌بهره‌ای اخیر بنگاه‌ها ورنش
doc number: 5709 -> title: https://www.farsnews.ir/news/14001010000506 اخبار: ر.ه.ای: استکل/14001010000506 حق صون کینه استیضایی درخصوصی فوق‌العاده: به بی‌بهره‌ای اخیر بنگاه‌ها ورنش
doc number: 6108 -> title: https://www.farsnews.ir/news/14001010000506 اخبار: ر.ه.ای: استکل/14001010000506 حق صون کینه استیضایی درخصوصی فوق‌العاده: به بی‌بهره‌ای اخیر بنگاه‌ها ورنش
doc number: 5657 -> title: https://www.farsnews.ir/news/14001010000506 اخبار: ر.ه.ای: استکل/14001010000506 حق صون کینه استیضایی درخصوصی فوق‌العاده: به بی‌بهره‌ای اخیر بنگاه‌ها ورنش
doc number: 3141 -> title: https://www.farsnews.ir/news/140011030003201 اخبار: ر.ه.ای: استکل/140011030003201 حق صون کینه استیضایی درخصوصی فوق‌العاده: به بی‌بهره‌ای اخیر بنگاه‌ها ورنش
doc number: 5294 -> title: https://www.farsnews.ir/news/1400101030003201 اخبار: ر.ه.ای: استکل/1400101030003201 حق صون کینه استیضایی درخصوصی فوق‌العاده: به بی‌بهره‌ای اخیر بنگاه‌ها ورنش
doc number: 4587 -> title: https://www.farsnews.ir/news/1400101030003201 اخبار: ر.ه.ای: استکل/1400101030003201 حق صون کینه استیضایی درخصوصی فوق‌العاده: به بی‌بهره‌ای اخیر بنگاه‌ها ورنش
doc number: 6744 -> title: https://www.farsnews.ir/news/14000926000189 اخبار: ر.ه.ای: استکل/14000926000189 حق صون کینه استیضایی درخصوصی فوق‌العاده: به بی‌بهره‌ای اخیر بنگاه‌ها ورنش

```

▶ `champion_similarity(query)`

```
<ipython-input-55-97d8424d142a>: RuntimeWarning: invalid value encountered in double_scalars
score=temp/(np.linalg.norm(list(query_tf(query).values()))*norm2_doc(query, i))

doc number: 3057 -> title: باشگاه گلپور: با افتتاح کسپنه بین‌الملل تاریخی‌ترین استاد رزم خورد/مادام پتوسی را فراموش کردید؟
doc number: 3457 -> title: عزنزلیختم: ما را به جانوگری می‌تواند کند برکت است/ هژوی از کیروش بیگانه/مهدی/اسک اسوجیج چه قست؟
doc number: 3141 -> title: کسپنه استینفد برای تروک‌های گلی گلر از تلید کرد/ استیمیزا به تیم سیرانی برگردانده می‌شود
doc number: 6764 -> title: اعلام اثر کسپنه اضافی در خصوص بازیکن گلی گلر در بازی با یکتا، اسکان و سیاهان بازیانه شد
doc number: 6128 -> title: کسپنه استینفد برای تروک‌های گلی گلر از تلید کرد/ استیمیزا به تیم سیرانی برگردانده می‌شود
doc number: 4918 -> title: عزنزلی ختم: قوتیل ایران از وضعیت اسفیلر گشتنه ضرر کرد
doc number: 1950 -> title: درفوشان/ رئیس فدراسیون فوتبال از کسپنه با جرمیجی برکتان شد/اذا خست بی‌پنده‌های شیطانی پری می‌شود
doc number: 2268 -> title: عرمشای: تصمیم گیری رتبه‌ای برای بازیکنان عزنزلی ختم جیتیجی بود/ تعلیبت عیضی اخلاقی غرضوتی است
doc number: 6323 -> title: اقم برنشتامی ولی اجازه ندانند VAR عزنزلی ختم: وقتی به فراسونین اقم یک خشک دم در آن نبود/ اری
doc number: 5102 -> title: کسپنه: بنوی عمل ختم-ما از آسیا بود/ کسپنه استکلان: اقم کسپنه بنوی عمل ختم-ما از آسیا بود
doc number: 139 -> title: کسپنه: بنوی عمل ختم-ما از آسیا بود/ کسپنه استکلان: اقم کسپنه بنوی عمل ختم-ما از آسیا بود
doc number: 5583 -> title: واکش نفت ابدان به شکایت یکتا از سکی با ماجای به مدیخلل کسپنه تهرانی و فراسونین اقم
doc number: 4603 -> title: اخلای/ عزنزلی ختم: اقم کسپنه بنوی عمل ختم-ما از آسیا بود/ کسپنه استکلان: اقم کسپنه بنوی عمل ختم-ما از آسیا بود
doc number: 1824 -> title: اخلای/ عزنزلی ختم: اقم کسپنه بنوی عمل ختم-ما از آسیا بود/ کسپنه استکلان: اقم کسپنه بنوی عمل ختم-ما از آسیا بود
doc number: 2626 -> title: اخلای/ عزنزلی ختم: اقم کسپنه بنوی عمل ختم-ما از آسیا بود/ کسپنه استکلان: اقم کسپنه بنوی عمل ختم-ما از آسیا بود
doc number: 5252 -> title: اخلای/ عزنزلی ختم: اقم کسپنه بنوی عمل ختم-ما از آسیا بود/ کسپنه استکلان: اقم کسپنه بنوی عمل ختم-ما از آسیا بود
doc number: 7305 -> title: اخلای/ عزنزلی ختم: اقم کسپنه بنوی عمل ختم-ما از آسیا بود/ کسپنه استکلان: اقم کسپنه بنوی عمل ختم-ما از آسیا بود
doc number: 5221 -> title: اخلای/ عزنزلی ختم: اقم کسپنه بنوی عمل ختم-ما از آسیا بود/ کسپنه استکلان: اقم کسپنه بنوی عمل ختم-ما از آسیا بود
doc number: 5005 -> title: اخلای/ عزنزلی ختم: اقم کسپنه بنوی عمل ختم-ما از آسیا بود/ کسپنه استکلان: اقم کسپنه بنوی عمل ختم-ما از آسیا بود
doc number: 1974 -> title: اخلای/ عزنزلی ختم: اقم کسپنه بنوی عمل ختم-ما از آسیا بود/ کسپنه استکلان: اقم کسپنه بنوی عمل ختم-ما از آسیا بود
```



۲) موارد ب و ت را با روش مکانی فاز یک نیز تکرار کنید و نتایج دو حالت را با هم مقایسه و تحلیل کنید.

با توجه به اینکه در بخش اول با توجه به وجود یا عدم وجود یک کلمه از کوئری در اسناد، اسناد منتخب را انتخاب می‌کنیم در نتیجه هنگامی که کوئری های بالا در ” “ قرار دهیم در این صورت اسنادی را به ما برمی‌گرداند که تنها شامل کوئری می‌هستند که ترتیب آن هم رعایت شده است در صورتی که در بخش دوم معیار انتخاب ما تعداد تکرار تک کلمات ما در کوئری ورودی است که در نهایت با در حساب کردن tf.idf برای اسناد مختلف و در نهایت محاسبه شباهت کسینوسی اسناد مشابه با کوئری مشتابه‌ترین اسناد را به کوئری ورودی انتخاب می‌کنیم و در خروجی قرار می‌دهیم.

ب) اسنادی که کلمه «فدراسیون فوتبال ایران» را به همین ترتیب حداقل یکبار در خود دارند.

مشاهده می‌شود که این اسناد دقیق‌تر هستند و بهتر می‌توانن نیاز کاربری که یک کوئری دقیق می‌خواهد را برطرف کند.

```
multi_word_query(query, must_query)

doc number: 3072 -> title: url: https://www.farsnews.ir/news/14001113000262/
doc number: 3073 -> title: url: https://www.farsnews.ir/news/14001113000246/
doc number: 5122 -> title: url: https://www.farsnews.ir/news/14001019000156/
doc number: 3077 -> title: url: https://www.farsnews.ir/news/14001113000182/
doc number: 3080 -> title: url: https://www.farsnews.ir/news/14001113000131/
doc number: 511 -> title: url: https://www.farsnews.ir/news/14001217000221/
doc number: 1548 -> title: url: https://www.farsnews.ir/news/14001217000221/
doc number: 4114 -> title: url: https://www.farsnews.ir/news/14001030000639/
doc number: 6170 -> title: url: https://www.farsnews.ir/news/14001004000429/
doc number: 3101 -> title: url: https://www.farsnews.ir/news/14001126000353/
doc number: 2078 -> title: url: https://www.farsnews.ir/news/14001126000763/
doc number: 6174 -> title: url: https://www.farsnews.ir/news/14001126000763/
doc number: 3615 -> title: url: https://www.farsnews.ir/news/14001106000531/
doc number: 2087 -> title: url: https://www.farsnews.ir/news/14001126000407/
doc number: 2088 -> title: url: https://www.farsnews.ir/news/14001126000353/
doc number: 6697 -> title: url: https://www.farsnews.ir/news/14001126000353/
doc number: 6191 -> title: url: https://www.farsnews.ir/news/14001004000109/
doc number: 6704 -> title: url: https://www.farsnews.ir/news/14000927000142/
doc number: 6193 -> title: url: https://www.farsnews.ir/news/14001004000049/
doc number: 2098 -> title: url: https://www.farsnews.ir/news/14001004000049/
```

ت) اسنادی که کلمه «فدراسیون فوتبال گابن» را به همین ترتیب حداقل یکبار در خود دارند.

همانطور که مشاهده می‌شود این اسناد عبارت دقیق را در خود دارند که به همین دلیل از تعداد کمی هم برخوردار هستند.

```
[152] def print_result(final):
      for i in range(len(final)):
          print("doc number: {} -> title: {} url: {}".format(df.index[final[i]], df.title[final[i]], df.url[final[i]]))

multi_word_query(query, must_query)

doc number: 6624 -> title: url: https://www.farsnews.ir/news/14000928000235/
doc number: 3141 -> title: url: https://www.farsnews.ir/news/14001112000
doc number: 6764 -> title: url: https://www.farsnews.ir/news/14000925/
doc number: 5709 -> title: url: https://www.farsnews.ir/news/14001010000508/
doc number: 820 -> title: url: https://www.farsnews.ir/news/14001202000284/
doc number: 1685 -> title: url: https://www.farsnews.ir/news/14001202000284/
doc number: 3101 -> title: url: https://www.farsnews.ir/news/14001126000353/
```