گزارش پروژه بازیابی اطلاعات

فاز1 : بخش 3 – پاسخ دهی به پرسمان کاربر

ابتدا کد را توضیح داده سپس تست های لازم را میگیریم. در این بخش هم به فایل حاوی اطلاعات شاخص مکانی و هم به فایل اصلی حاوی اخبار نیاز داریم پس هردورا فراخوانی میکنیم و فایل اصلی را در ساختاری که در بخش اول تعریف کرده بودیم میریزیم به صورتی که دسترسی به تیتر و آدرس خبرها آسان تر باشد:

def openFiles():  
 global data,positional\_index  
 # Opening positional index file  
 file\_path = 'C:/Users/Samin/Desktop/University/Term 7/Information Retrieval/Project/Data/IR\_data\_news\_12k\_positional\_index\_dic.json'  
 try:  
 with open(file\_path, 'r', encoding='utf-8') as f:  
 positional\_index = json.load(f)  
 print("Positional Index File opened successfully!")  
 except IOError:  
 print("Error opening file.")  
  
 # Opening origin file  
 file\_path = 'C:/Users/Samin/Desktop/University/Term 7/Information Retrieval/Project/Data/IR\_data\_news\_12k.json'  
 try:  
 with open(file\_path, 'r', encoding='utf-8') as f:  
 data\_raw = json.load(f)  
 print("Origin File opened successfully!")  
 except IOError:  
 print("Error opening file.")  
  
 for docID, body in data\_raw.items():  
 data[docID] = {}  
 data[docID]['title'] = body['title']  
 data[docID]['content'] = body['content']  
 data[docID]['url'] = body['url']

پس از دریافت کوئری از کاربر این تابع فراخوانی میشود، سپس کوئری ورودی به فانکشن دیگری فرستاده خواهد که در ادامه توضیح میدهیم:

def main():  
 inputQ = input('Enter Query : ')  
   
 openFiles()  
 queryProcessor(inputQ)

تابع queryProcessor کوئری ورودی را گرفته و در قدم اول لیست عبارات، کلماتی که مدنظر هستند و کلماتی که مدنظر نیستند را از آن استخراج میکند:

def queryProcessor(query):  
 global data,positional\_index  
 # call funcs on query  
 rejected\_terms = findRejected(query)  
 phrase\_terms = findPhrases(query)  
 actual\_terms = findWords(query)  
 print(f'Rejected list : {rejected\_terms}')  
 print(f'Phrase list : {phrase\_terms}')  
 print(f'Actual list : {actual\_terms}')

کلمات مدنظر از طریق این تابع استخراج میشوند:

def findWords(query):  
 # Remove words in quotes  
 query = re.sub(r'"([^"]+)"', "", query)  
  
 # Remove words starting with !  
 query = re.sub(r'!\w+\s?', "", query)  
 words = query.strip()  
 words\_list = []  
 pure\_words = re.sub(f'[{punctuation}؟،٪×÷»«]+', '', words)  
 preprocessed\_words = to\_stem(to\_remove\_stop\_words(to\_tokenize(to\_normalize(pure\_words))))  
 words\_list.append(preprocessed\_words)  
 return words\_list

همانطور که مشاهده میشود بخشی از کوئری که گیومه و علامت تعجب از آن جدا شده و بخش باقی شامل کلماتی خواهد بود که باید مستقلا در پاسخ حضور داشته باشند. لازم به ذکر است که این کلمات با همان توابع بخش اول فاز 1 که پیش پردازش را روی اخبار انجام داده بود پیش پردازش میشوند تا به همان ساختاری برسند که ذخیره سازی و ساخت دیکشنری روی آنها انجام شده بود.

عبارات توسط تابع زیر استخراج خواهند شد:

def findPhrases(query):  
 # finding phrases  
 phrases = []  
 phrases = re.findall('**\"**.\*?**\"**', query)  
 phrase\_dic = []  
 for phrase in phrases:  
 pure\_phrase = re.sub(f'[{punctuation}؟،٪×÷»«]+', '', phrase)  
 preprocessed\_phrase = to\_stem(to\_remove\_stop\_words(to\_tokenize(to\_normalize(pure\_phrase))))  
 phrase\_dic.append(preprocessed\_phrase)  
 return phrase\_dic

در این تابع عباراتی که در کوئری داخل گیومه هستند را گرفته پس از پیش پردازش در یک لیست ذخیره سازی میکنیم به صورتی که در هر سطر کلمات پیش پردازش شده یک عبارت وجود داشته باشد.

توسط تابع زیر کلماتی که مدنظر نیستند را استخراج میکنیم:

def findRejected(query):  
 # finding rejected terms  
 rejected\_terms = []  
 if '!' in query:  
 query\_parts = query.split()  
 for part in query\_parts:  
 if part.startswith('!'):  
 pure\_rejected = re.sub(f'[{punctuation}؟،٪×÷»«]+', '', part[1:])  
 preprocessed\_rejected = to\_stem(to\_remove\_stop\_words(to\_tokenize(to\_normalize(pure\_rejected))))  
 rejected\_terms.append(preprocessed\_rejected)  
 return rejected\_terms

همانطور که مشاهده میشود کلماتی که با ! شروع شده اند در لیست کلمات ممنوعه قرار خواهند گرفت و روی آنها پیش پردازش انجام خواهد شد.

حالا باید امتیاز دهی داک هارا مدیریت کنیم. در حالت اول شرایطی را مدیریت میکنیم که تنها با کلماتی کار داریم که مستقلا باید در متن حضور داشته باشند:

# scoring docs  
scores = {}  
# part 1 - finding actual terms  
for item in actual\_terms:  
 for term in item:  
 if term in positional\_index:  
 for docID, doc\_data in positional\_index[term].items():  
 if docID != 'total':  
 if docID not in scores:  
 scores[docID] = 5 # finding one accepted term  
 else:  
 scores[docID] += 5 # finding more than one accepted word

با توجه به لیستی از این کلمات که قبلا بدست اورده بودیم وجود انها را در شاخص مکانی بررسی کرده و داک هایی که شامل آنها هستند به لیست امتیازها اضافه شده و امتیاز 5 میگیرند، اگر یک داک بیش از یک کلمه مدنظر را داشته باشد به امتیاز آن 5 تا اضافه خواهد شد.

در بخش بعدی مدیریت امتیاز دهی به عبارات را خواهیم داشت:

for phrase in phrase\_terms:  
 phrase\_check = {}  
 for term in phrase:  
 if term in positional\_index:  
 if term not in phrase\_check:  
 phrase\_check[term] = set()  
 keys = positional\_index[term].keys()  
 for key in keys:  
 if key != 'total':  
 phrase\_check[term].add(key)

برای هر عبارتی روی کلمات آن حلقه میزنیم و داک های شامل آنها را بصورت دیکشنری ای که کلید آن کلمه و مقدار آن ست داک های شامل آن است ذخیره میکنیم. در قدم بعدی اشتراک این ست هارا با توسط تابع اینترسکشن بدست می آوریم:

# Find the intersection of all documents containing each term in the phrase  
common\_docs = set.intersection(\*phrase\_check.values())

حالا باید چک کنیم در این داک هایی که حاوی این کلمات هستند فاصله کلمات به چه صورت است و آیا پشت سرهم آمده اند؟ برای این مطلب روی داک های مشترک حرکت کرده و برای هر ترم موقعیت آن را بدست می اوریم. متغیری تعریف میکنیم و از اولین کلمه به اندازه تعداد کلمات عبارت جلو میرویم و چک میکنیم آیا در موقعیت های بعدی از کلمه اول کلمات بعدی قرار داشته اند یا خیر

# Check if the terms in each common document are adjacent  
for doc\_id in common\_docs:  
 positions = {}  
 for term in phrase:  
 positions[term] = []  
 if doc\_id in positional\_index[term]:  
 positions[term].append(positional\_index[term][doc\_id]['positions'])  
 check = 1  
 for item in positions[phrase[0]]:  
 for position in item:  
 for i in range(1, len(phrase)):  
 p = position + i  
 for item2 in positions[phrase[i]]:  
 if p in item2:  
 check = check + 1

در چنین حلقه ای اگر عبارت مدنظر بیش از یک بار هم امده باشد چک میشود و درنظر گرفته میشود بنابراین به مقدار متغیر check اگر افزایش پیدا کرده بود به امتیاز آن داک می افزاییم:

if check > 1:  
 if doc\_id not in scores:  
 scores[doc\_id] = 0  
 scores[doc\_id] += check

حالا باید داک هایی که کلمات ممنوعه دارند را حذف کنیم، برای اینکار لیست کلمات ممنوعه را گرفته و داک ایدی شامل آن هارا ذخیره میکنیم، اگر این داک ایدی ها در لیست داک های امتیاز دهی شده حضور داشتند امتیازشان صفر خواهد شد.

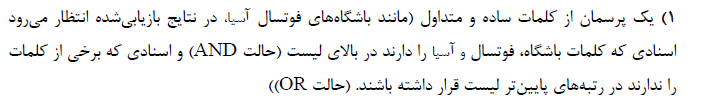
# part 3 - removing docs with rejected terms  
rejected\_docs = []  
for item in rejected\_terms:  
 for term in item:  
 if term in positional\_index:  
 keys = positional\_index[term].keys()  
 for key in keys:  
 if key != 'total' and key not in rejected\_docs:  
 rejected\_docs.append(key)  
 for docID in rejected\_docs:  
 if docID in scores:  
 del scores[docID] # remove docs with rejected words

نهایتا داک های امتیاز دهی شده سورت میشوند، اگر لیست داک های امتیاز دهی شده خالی بود یعنی داده ای پیدا نکرده ایم پس پیام مناسب را نشان میدهیم در غیر این صورت لیست داک های امتیاز دهی شده به تابعی برای نمایش ارسال خواهد شد:

sorted\_docs = sorted(scores.items(), key=lambda x: x[1], reverse=True)  
  
if len(sorted\_docs) > 0:  
 toPrint(sorted\_docs)  
else:  
 print('داده ای یافت نشد')

در این تابع اگر تعداد داک ها از 10 تا بیشتر بود 5 تای اول را به همراه تیتر و داک ایدی و ادرس و امتیازش نمایش میدهیم درغیر این صورت تمامی داک هارا نمایش میدهیم:

def toPrint(sorted\_docs):  
 global data  
 if len(sorted\_docs)>100:  
 for i in range(0, 5):  
 docID = sorted\_docs[i][0]  
 docScore = sorted\_docs[i][1]  
 title = data[docID]['title']  
 url = data[docID]['url']  
 print(f'{i + 1}. DocID = {docID} \n Title = {title} \n URL = {url}\n Score = {docScore}')  
 else:  
 for i in range(0, len(sorted\_docs)):  
 docID = sorted\_docs[i][0]  
 title = data[docID]['title']  
 docScore = sorted\_docs[i][1]  
 url = data[docID]['url']  
 print(f'{i + 1}. DocID = {docID} \n Title = {title} \n URL = {url}\n Score = {docScore}')



خروجی:

**Enter Query : باشگاه های فوتسال آسیا**

**Positional Index File opened successfully!**

**Origin File opened successfully!**

**Rejected list : []**

**Phrase list : []**

**Actual list : [['باشگاه', 'فوتسال', 'آسیا']]**

**1. DocID = 0**

**Title = اعلام زمان قرعه کشی جام باشگاه های فوتسال آسیا**

**URL = https://www.farsnews.ir/news/14001224001005/اعلام-زمان-قرعه-کشی-جام-باشگاه-های-فوتسال-آسیا**

**Score = 15**

**2. DocID = 797**

**Title = مدیرعامل پیکان: پتانسیل تهران برای میزبانی جام باشگاه‌ها بیشتر است/حافظیه در حد لیگ یک و دو هم نیست**

**URL = https://www.farsnews.ir/news/14001214000307/مدیرعامل-پیکان-پتانسیل-تهران-برای-میزبانی-جام-باشگاه‌ها-بیشتر-است**

**Score = 15**

**3. DocID = 1306**

**Title = تشریح جزئیات دریافت مجوز حرفه‌ای برای تیم‌های فوتسال و تشکیل کمیته ویژه در هفته‌های پایانی لیگ**

**URL = https://www.farsnews.ir/news/14001207000286/تشریح-جزئیات-دریافت-مجوز-حرفه‌ای-برای-تیم‌های-فوتسال-و-تشکیل-کمیته**

**Score = 15**

**4. DocID = 1535**

**Title = چرا تیم ملی فوتسال به تایلند اعزام نمی شود؟**

**URL = https://www.farsnews.ir/news/14001203001189/چرا-تیم-ملی-فوتسال-به-تایلند-اعزام-نمی-شود**

**Score = 15**

**5. DocID = 1671**

**Title = مجوز حرفه‌ای AFC از فوتبال به فوتسال رسید/باشگاه‌های ایران به هوش باشند**

**URL = https://www.farsnews.ir/news/14001202000526/مجوز-حرفه‌ای-AFC-از-فوتبال-به-فوتسال-رسید-باشگاه‌های-ایران-به-هوش**

**Score = 15**

**6. DocID = 1824**

**Title = ساکت: فراز و نشیب‌های سپاهان برای ما هم تلخ بود/ نویدکیا مورد حمایت باشگاه است**

**URL = https://www.farsnews.ir/news/14001130000599/ساکت-فراز-و-نشیب‌های-سپاهان-برای-ما-هم-تلخ-بود-نویدکیا-مورد-حمایت**

**Score = 15**

**7. DocID = 3457**

**Title = عزیزی‌خادم:‌ ما را به جادوگری متهم کردند ولی جادوگری شرک است/هنوز برای کی‌‌روش مالیات می‌دهیم/مگر اسکوچیچ رفته است؟**

**URL = https://www.farsnews.ir/news/14001108000181/عزیزی‌خادم-‌-ما-را-به-جادوگری-متهم-کردند-ولی-جادوگری-شرک-است-هنوز-برای**

**Score = 15**

**8. DocID = 4213**

**Title = حسن زاده: فدراسیون فوتبال شمسایی و تیم ملی را حمایت کند/ فعلا قصد خداحافظی ندارم**

**URL = https://www.farsnews.ir/news/14001029000320/حسن-زاده-فدراسیون-فوتبال-شمسایی-و-تیم-ملی-را-حمایت-کند-فعلا-قصد**

**Score = 15**

**9. DocID = 4562**

**Title = صانعی: از اعتبار فوتسال ایران دفاع می کنیم**

**URL = https://www.farsnews.ir/news/14001025000814/صانعی-از-اعتبار-فوتسال-ایران-دفاع-می-کنیم**

**Score = 15**

**10. DocID = 4671**

**Title = سفال منش: حضور شمسایی نقطه عطف در فوتسال است/رزومه 3 مربی اسپانیایی را به فدراسیون ارسال کردیم**

**URL = https://www.farsnews.ir/news/14001023000871/سفال-منش-حضور-شمسایی-نقطه-عطف-در-فوتسال-است-رزومه-3-مربی-اسپانیایی-را**

**Score = 15**

**11. DocID = 5350**

**Title = مربی ایرانی گزینه سرمربیگری تیم ملی فوتسال اندونزی شد**

**URL = https://www.farsnews.ir/news/14001015000897/مربی-ایرانی-گزینه-سرمربیگری-تیم-ملی-فوتسال-اندونزی-شد**

**Score = 15**

**12. DocID = 6147**

**Title = درخشش بانوی فوتسالیست در مسابقات بین المللی /آخوندی:حجاب در فوتبال و فوتسال بانوان آسیا در حال گسترش است**

**URL = https://www.farsnews.ir/news/14001004000725/درخشش-بانوی-فوتسالیست-در-مسابقات-بین-المللی-آخوندی-حجاب-در-فوتبال-و**

**Score = 15**

**13. DocID = 6370**

**Title = عزیزی‌خادم:اسکوچیچ با قدرت به کارش ادامه می‌دهد/داوران ما VAR را آموزش نبینند از چرخه حذف می‌شوند**

**URL = https://www.farsnews.ir/news/14001001000441/عزیزی‌خادم-اسکوچیچ-با-قدرت-به-کارش-ادامه-می‌دهد-داوران-ما-VAR-را-آموزش**

**Score = 15**

**14. DocID = 8**

**Title = مدیر بی‌دستاورد، رئیس شد/ روزگار تلخ‌تر از تلخ برای والیبال ارومیه؟**

**URL = https://www.farsnews.ir/news/14001224000811/مدیر-بی‌دستاورد-رئیس-شد-روزگار-تلخ‌تر-از-تلخ-برای-والیبال-ارومیه**

**Score = 10**

**15. DocID = 29**

**Title = مدیرعامل آلومینیوم: مرفاوی می‌داند هیچ باشگاهی با یک نفر صحبت نمی‌کند/استقلالی‌ها لیگ را گرفته‌اند**

**URL = https://www.farsnews.ir/news/14001224000282/مدیرعامل-آلومینیوم-مرفاوی-می‌داند-هیچ-باشگاهی-با-یک-نفر-صحبت-نمی‌کند**

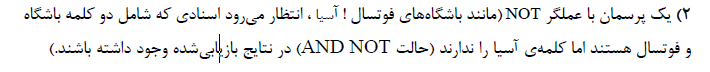
**Score = 10**

همانطور که مشاهده میشود تمامی کلمات کوئری مستقلا شناسایی و بررسی شده اند و پس از امتیاز دهی 5 داک برتر نمایش داده شده اند. تعداد داک های یافت شده را برای نمایش بهتر به 15 تا افزایش دادیم و مشاهده میکنیم که امتیاز آنها در داک های اول 15 یعنی هر سه کلمه یافت شده و شرایط and برقرار بوده ولی در داک های بعدی به امتیاز 10 هم رسیدیم یعنی تنها دو کلمه یافت شده اند که شرایط or برقرار است.

برای مثال متن اولین داک یافت شده را بررسی میکنیم:



همانطور که مشاهده میشود این خبر حاوی کوئری مدنظر ما بوده و این کوئری چندین بار هم تکرار شده است.



خروجی:

**Enter Query : باشگاه های فوتسال !آسیا**

**Positional Index File opened successfully!**

**Origin File opened successfully!**

**Rejected list : [['آسیا']]**

**Phrase list : []**

**Actual list : [['باشگاه', 'فوتسال']]**

**1. DocID = 81**

**Title = ماجدی: فوتبال کشور به تغییرات نیاز دارد**

**URL = https://www.farsnews.ir/news/14001223000539/ماجدی-فوتبال-کشور-به-تغییرات-نیاز-دارد**

**Score = 10**

**2. DocID = 426**

**Title = اتفاق عجیب در لیگ نوجوانان/ قهرمانی سایپا پس گرفته شد!+نامه**

**URL = https://www.farsnews.ir/news/14001218000265/اتفاق-عجیب-در-لیگ-نوجوانان-قهرمانی-سایپا-پس-گرفته-شد-نامه**

**Score = 10**

**3. DocID = 453**

**Title = دبیر هیات فوتبال اصفهان: ابلاغی برای حضور تماشاگران نداشتیم/امیدوارم روزی مساوات رعایت شود**

**URL = https://www.farsnews.ir/news/14001217000981/دبیر-هیات-فوتبال-اصفهان-ابلاغی-برای-حضور-تماشاگران-نداشتیم-امیدوارم**

**Score = 10**

**4. DocID = 467**

**Title = سرمربی شهرخودرو جریمه شد/محرومیت برای بازیکن فوتسال**

**URL = https://www.farsnews.ir/news/14001217000931/سرمربی-شهرخودرو-جریمه-شد-محرومیت-برای-بازیکن-فوتسال**

**Score = 10**

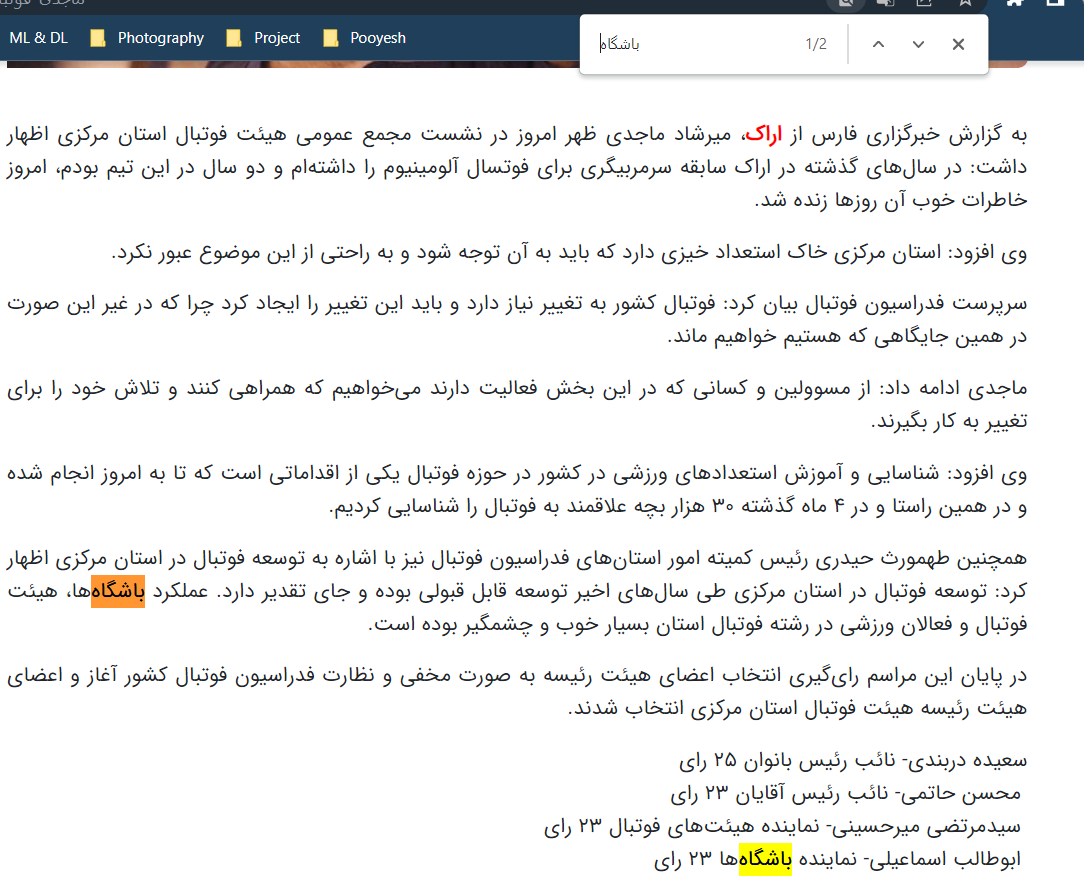
**5. DocID = 493**

**Title = اوج حساسیت در هفته آخر لیگ برتر فوتسال/ انصراف شهروند از بازی مقابل مس**

**URL = https://www.farsnews.ir/news/14001217000534/اوج-حساسیت-در-هفته-آخر-لیگ-برتر-فوتسال-انصراف-شهروند-از-بازی-مقابل-مس**

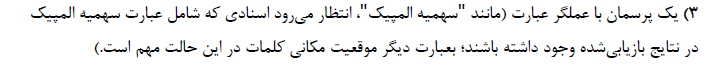
**Score = 10**

همانطور که مشاهده میشود آسیا به عنوان کلمه ممنوعه شناسایی شده و تنها داک هایی پیدا شده اند که دو کلمه باشگاه و فوتسال را داشته و آسیا را نداشته اند.





و برای آسیا کلمه ای پیدا نمیشود.



خروجی:

**Enter Query : "سهمیه المپیک"**

**Positional Index File opened successfully!**

**Origin File opened successfully!**

**Rejected list : []**

**Phrase list : [['سهمیه', 'المپیک']]**

**Actual list : [[]]**

**1. DocID = 228**

**Title = گلخندان: مدال فروغی در المپیک مشوق خوبی برای همه بود/ نمی‌توان برای بازی‌های آسیایی قول داد**

**URL = https://www.farsnews.ir/news/14001221000090/گلخندان-مدال-فروغی-در-المپیک-مشوق-خوبی-برای-همه-بود-نمی‌توان-برای**

**Score = 4**

**2. DocID = 3907**

**Title = تیراندازان در کدام رویدادها المپیکی می‌شوند؟**

**URL = https://www.farsnews.ir/news/14001103000111/تیراندازان-در-کدام-رویدادها-المپیکی-می‌شوند**

**Score = 4**

**3. DocID = 5862**

**Title = سهرابیان: قایق‌ها 21 ماه در گمرک اسیر شد/ممکن بود راضی شویم قایق‌ها را در چهارشنبه سوری بسوزانند**

**URL = https://www.farsnews.ir/news/14001008000645/سهرابیان-قایق‌ها-2-ماه-در-گمرک-اسیر-شد-ممکن-بود-راضی-شویم-قایق‌ها-را**

**Score = 3**

**4. DocID = 164**

**Title = رستمیان: رنکینگ ما در کسب سهمیه المپیک مهم است/حضور در جام جهانی قطعی نیست**

**URL = https://www.farsnews.ir/news/14001221001088/رستمیان-رنکینگ-ما-در-کسب-سهمیه-المپیک-مهم-است-حضور-در-جام-جهانی-قطعی**

**Score = 3**

**5. DocID = 4836**

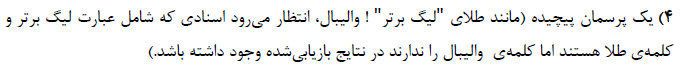
**Title = صید نماینده ایران در بازی‌های المپیک زمستانی شد**

**URL = https://www.farsnews.ir/news/14001022000538/صید-نماینده-ایران-در-بازی‌های-المپیک-زمستانی-شد**

**Score = 3**

همانطور که مشاهده میشود داک هایی برگردانده شده اند که این عبارت را داشته اند و با توجه به تعداد مرتبه هایی که این عبارت در داک ها امده امتیاز دهی انجام شده است.





خروجی:

**Enter Query : طلای "لیگ برتر" !والیبال**

**Positional Index File opened successfully!**

**Origin File opened successfully!**

**Rejected list : [['والیبال']]**

**Phrase list : [['لیگ', 'بر']]**

**Actual list : [['طلا']]**

**1. DocID = 5632**

**Title = هفته سیزدهم لیگ برتر فوتبال| تداوم شکست ناپذیری استقلال در اراک؟/ نبرد قرمزپوشان تهران و تبریز**

**URL = https://www.farsnews.ir/news/14001011000035/هفته-سیزدهم-لیگ-برتر-فوتبال|-تداوم-شکست-ناپذیری-استقلال-در-اراک-نبرد**

**Score = 17**

**2. DocID = 5992**

**Title = هفته دوازدهم لیگ برتر|سپاهان - پرسپولیس؛ نبرد آسیایی در زمین بیطرف/استقلال مقابل مانع فولادی و دربی کرمان**

**URL = https://www.farsnews.ir/news/14001006000037/هفته-دوازدهم-لیگ-برتر|سپاهان--پرسپولیس-نبرد-آسیایی-در-زمین-بیطرف**

**Score = 16**

**3. DocID = 820**

**Title = هفته بیست و یکم لیگ برتر|منصوریان به دنبال عاقبت بخیری با پرسپولیس/تقابل استقلال و قلعه‌نویی بدون بازیکن گابنی**

**URL = https://www.farsnews.ir/news/14001211000992/هفته-بیست-و-یکم-لیگ-برتر|منصوریان-به-دنبال-عاقبت-بخیری-با-پرسپولیس**

**Score = 15**

**4. DocID = 2014**

**Title = هفته هجدهم لیگ برتر| پرسپولیس- آلومینیوم؛ شاگردان گل‌محمدی به دنبال انتقام/استقلال در تبریز شکست ناپذیر می‌ماند؟**

**URL = https://www.farsnews.ir/news/14001127000107/هفته-هجدهم-لیگ-برتر|-پرسپولیس-آلومینیوم-شاگردان-گل‌محمدی-به-دنبال**

**Score = 13**

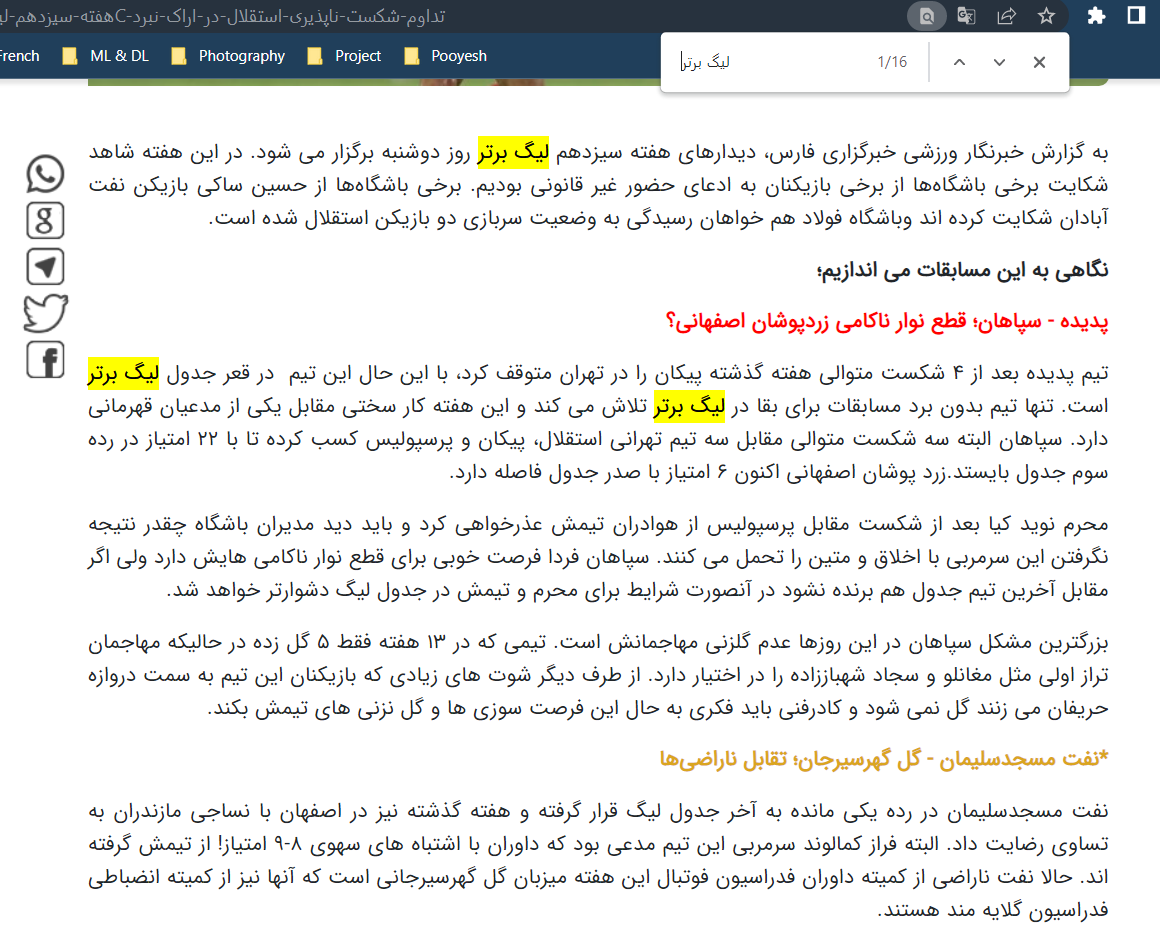
**5. DocID = 974**

**Title = روایتی جالب از تیم قهرمان لیگ برتر واترپلو/ به نام شهیدی که تیم ملی و مدال را رها کرد و رفت!**

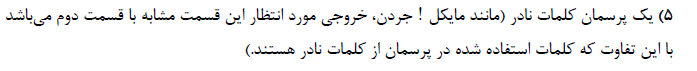
**URL = https://www.farsnews.ir/news/14001208000567/روایتی-جالب-از-تیم-قهرمان-لیگ-برتر-واترپلو-به-نام-شهیدی-که-تیم-ملی-و**

**Score = 12**

همانطور که مشاهده میشود کلمه طلا مستقلا درنظر گرفته شده، عبارت لیگ برتر باهم بررسی شده و کلمه والیبال ممنوعه بوده و بنابر همین روند 5 داک اول نمایش داده شده اند.

و برای کلمه آسیا چیزی پیدا نمیشود.



خروجی:

**Enter Query : مایکل !جردن**

**Positional Index File opened successfully!**

**Origin File opened successfully!**

**Rejected list : [['جردن']]**

**Phrase list : []**

**Actual list : [['مایکل']]**

**1. DocID = 722**

**Title = اتومبیل راننده اسطوره جهان به حراج گذاشته می‌شود**

**URL = https://www.farsnews.ir/news/14001215000107/اتومبیل-راننده-اسطوره-جهان-به-حراج-گذاشته-می‌شود**

**Score = 5**

**2. DocID = 2636**

**Title = تیم منتخب هفته قطر بدون ایرانی‌ها/3 بازیکن خطرناک رقیب سپاهان را بشناسید+فیلم**

**URL = https://www.farsnews.ir/news/14001119000103/تیم-منتخب-هفته-قطر-بدون-ایرانی‌ها-3-بازیکن-خطرناک-رقیب-سپاهان-را**

**Score = 5**

**3. DocID = 3774**

**Title = کار سخت مدافعان تیم ملی مقابل دومین گلزن برتر لیگ ستارگان در آزادی**

**URL = https://www.farsnews.ir/news/14001104000540/کار-سخت-مدافعان-تیم-ملی-مقابل-دومین-گلزن-برتر-لیگ-ستارگان-در-آزادی**

**Score = 5**

**4. DocID = 4085**

**Title = دومین خط حمله زهردار قطر منتظر سپاهان/رقابت جذاب ابراهیمی، استراماچونی و اسماعیلی+عکس**

**URL = https://www.farsnews.ir/news/14001101000113/دومین-خط-حمله-زهردار-قطر-منتظر-سپاهان-رقابت-جذاب-ابراهیمی-استراماچونی**

**Score = 5**

**5. DocID = 6122**

**Title = تبعات یک شکست عجیب؛ همیلتون آتش به مالش زد + عکس**

**URL = https://www.farsnews.ir/news/14001005000088/تبعات-یک-شکست-عجیب-همیلتون-آتش-به-مالش-زد--عکس**

**Score = 5**

مشابه بخش دوم پاسخ میگیریم:





در تمامی بخش ها اخبار با تیتر و آدرس و متن آورده شده و مشاهده میشود که کوئری وارد شده با محتوای خبر مرتبط بوده است.